EXHIBIT A

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Robert James, et al.	**Examiner:**	Juliet C. Switzer
Serial No.:	10/800,322	**Art Unit:**	1634
Filed:	March 12, 2004	**Docket:**	17530
For:	NUCLEIC ACID MARKERS FOR USE IN DETERMINING PREDISPOSITION TO NEOPLASM AND/OR ADENOMA	**Confirmation No.:**	2289

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF DR. SUSANNE PEDERSEN UNDER 37 C.F.R. §1.132

Sir:

I, Susanne Pedersen, hereby declare as follows:

1. I am the Chief Scientific Officer at Clinical Genomics Pty Ltd, CG (Sydney, Australia) and am responsible for managing CG research and development. I joined Clinical Genomics in February 2007 after six years as a Senior Research Scientist and Project Leader at Proteome Systems Ltd (now known as Tyrian's Diagnostics, Sydney) where I managed biomarker discovery in respiratory diseases, autoimmune disorders and cancer. I am the inventor of several biomarker-discovery method patents.

2. I hold a Master of Science (MSc) degree and a Doctorate Degree in Molecular Biology from Southern University of Denmark (Odense, Denmark) and The Royal National Hospital (Copenhagen, Denmark). I have authored a number of scientific publications, including publications relating to biomarker discovery and biomarker method development. A true and correct copy of my curriculum vitae is attached hereto as **Exhibit 1.**

3. I have reviewed the above-identified application (the '322 application), and am familiar with the subject matter disclosed therein. I understand that the subject matter presently claimed in the '322 application is directed to, *inter alia*, a method for determining the onset of colorectal adenoma in a human by measuring the level of expression, in a blood, serum, stool or gastrointestinal tract sample, of a nucleic acid molecule that contains the nucleotide sequence as set forth in SEQ ID NO: 7, or a nucleic acid molecule that contains a nucleotide sequence complementary to SEQ ID NO: 7.

4. I have read the Office Action dated February 26, 2009 issued in the '322 application, and I have been asked to provide comments on issues raised in the Office Action.

Relationship between SEQ ID NO: 7 and KIAA1199

5. The Examiner stated in the Action that she was not able to establish a relationship between instant SEQ ID NO: 7 and the sequence set forth in GenBank database under accession No. AB033025.

6. KIAA1199 is the designation of the genomic sequence available in the NCBI GenBank database under accession number AB033025. The NCBI RefSeq annotation for the KIAA1199 gene also provides one representative mRNA transcript of KIAA1199. Based on this annotation and using publicly available bioinformatics tools including for example BLAST or ClustalW for alignment of AB033025 and SEQ ID NO: 7, it is my determination that SEQ ID NO: 7 lies between the exon 1 and exon 2 of NCBI RefSeq KIAA1199 gene locus. Therefore, if one were to align SEQ ID NO: 7 with the mRNA sequence of KIAA1199 (which is apparently what the Examiner did), one would not find any similarity between the two sequences.

7. While the NCBI RefSeq database only annotates one representative transcript transcribed from the KIAA1199A gene locus, the AceView database, which is the NCBI EST/cDNA database also available in the art, indicates the existence of other splice variant forms (i.e. mRNA) transcribed from the KIAA1199A gene. In reviewing the AceView

database, the KIAA1199 splice variants "j-u", "h-u", "i-u" and "g-u" all represent experimentally identified cDNA clones transcribed from so-called "intronic" sequences from the RefSeq annotated KIAA1199 gene. See **Exhibit 2**. SEQ ID NO: 7 is simply another example of such transcripts.

8. Further, the RefSeqGene annotation has various weakness, as acknowledged by the NCBI web site, **http://www.ncbi.nlm.nih.gov/projects/RefSeq/RSG/**:

> "RefSeq, a subset of NCBI's Reference Sequence (RefSeq) project, defines genomic sequences of well-characterized genes to be used as reference standards. These sequences, labeled with the keyword RefSeqGene, serve as a stable foundation for reporting mutations, for establishing conventions for numbering exons and introns, and for defining the coordinates of other biologically significant variation. RefSeq mRNA and protein sequences already support these functions, but have the obvious weakness of not providing explicit coordinates for flanking or intronic sequence. RefSeq chromosome sequences also support these functions, but have awkwardly large coordinate values that will change when the sequence is updated with a new genome build. Sequences of the RefSeqGene project will counter both of these drawbacks by providing gene-specific genomic sequence for each gene, as well as including upstream and downstream flanking regions. If modifications must be made to any RefSeqGene sequence, it will be versioned and tools will be provided to facilitate conversion of coordinates. The RefSeqGene sequences will also be placed on the reference chromosome, and current chromosome coordinates will be available because of that re-alignment." (Emphasis added.)

9. The following additional experimental data in **Exhibit 3** are provided to show that SEQ ID NO:7 is in fact a transcript of the KIAA1199 gene. The experiments were either performed by myself or under my direct supervision, and data interpretation and resulting reports were conducted by me.

(i) Figure 1 illustrates the differential expression across the 29 exons of KIAA1199 as designated in NCBI GenBank by measuring the hybridization of RNA extracted from colon specimens from 30 normal subjects and 19 subjects with either adenomas or colorectal cancer. All probe sets across the 29 exons with the singe exception of exon 18 show upregulation of RNA derived from the KIAA1199 gene locus. For example, RNA samples derived from exon

1 and exon 2, which flank SEQ ID NO: 7, show up-regulation in colorectal neoplasia. In contrast, the flanking genes to KIAA1199 on the (+) strand, MESDC1 and FAM108C1, were not differentially expressed, Figure 2.

(ii) Figure 3 illustrates the generation of a dominant PCR product using a forward primer in the 3'-end of SEQ ID NO: 7 and a reverse primer in the 5'-end of NCBI GenBank exon 2 using an end-point PCR based analysis with RNA extracted from colon tissue specimens from 2 normal subjects and 2 subjects with adenomas or cancers. These data support the notion that SEQ ID NO: 7 is an RNA transcript derived from the KIAA1199 gene locus.

10. Clearly, SEQ ID NO: 7 is an integral part of the genomic sequence of KIAA1199, and KIAA1199 is transcribed with the SEQ ID NO: 7 sequence forming part of the mRNA.

Sample Source

11. The Examiner also raised the issue that the '322 application only teaches overexpression of SEQ ID NO: 7 in colorectal tissue biopsy samples, and experimentation and data for fecal and blood samples are absent.

12. Experiments were conducted under my supervision, which demonstrate upregulation in the level of expression of KIAA1199 in stool samples and in serum samples. Specifically, the results in **Exhibit 4,** Part 1, show that increased levels of KIAA1199 protein in stool samples were detected an indirect ELISA using a monoclonal antibody directed to KIAA1199 protein. In my opinion, increased level of KIAA1199 mRNA transcripts would also have occurred in these stool samples as well.

13. It is known that neoplastic colorectal epithelial cells may be exfoliated in the gut lumen and concomitantly the stools, where they are detectable. Leakage, escape and migration (sometimes involving apoptosis and anoikis) of neoplastic cells into the lymphatic or peripheral

blood vessel system are well documented in the literature. This phenomenon is a characteristic of metastasis in which neoplastic cells spread from original organ to a secondary, perhaps distant, organ via the vasculature. Appearance of colorectal neoplastic cells in the blood and detection of their cellular content (e.g., RNA, protein and DNA) has been documented in the literature. For example, Galamb et al. (2008) (also provided in **Exhibit 4**) demonstrated the fact that colorectal tissue RNA markers can be found in the peripheral blood. Further, the results in **Exhibit 4**, Part 2, show that approximately 30% of adenoma patients exhibited a significant increase in KIAA1199 mRNA in the plasma.

14. It is noted that the techniques employed in the experiments shown in Exhibit 4 were all available to those skilled in the art when the '322 application was first filed. Therefore, it is my opinion that once an upregulated expression of a biomarker is established based on tissue biopsy sample, the experimentation involved in confirming that elevated expression can also be detected in stool and blood samples would be routine and not excessive.

Determination of an increase in expression

15. The '322 application discloses that clones 8-2d and 12-2f, to which SEQ ID NO: 7 corresponds, were up-regulated by 50 and 45 fold, respectively, in adenoma tissue samples. Additional experiments were performed under my supervision, which also show that there is a statistically significant increase in SEQ ID NO: 7 levels in adenoma patients as compared to normal patients. These additional validation data were presented as Exhibit 3 in the Response filed in the '322 application on December 1, 2008.

16. In **Exhibit 5** attached hereto, the original discovery data (provided in the '322 application) and the further validation data (submitted in the Response filed in the '322 application on December 1, 2008) are presented herein in Figures 4 and 5, respectively, in a graph format. It is clear from the graphs in Figures 4 and 5 that the mean level of expression of KIAA1199 in normal patients as opposed to adenoma patients is statistically significantly different. There is a statistically significant increase of expression of SEQ ID NO: 7 in each of the graphs. The Examiner should not be focusing on the few patients who appear above and

below the mean value, i.e. within the standard deviation, also referred to as the overlapping "whiskers" between the normal and adenoma data sets. This is equivalent to the ends of the bell curves shown in the previously presented bell diagram. It is the middle box which the Examiner should focus on since this represents the data median. The top and bottom of the box are the 25% and 75% interquartile range. The whiskers represent the minimum and maximum observations which are considered outliers. The phenotypes are differentially expressed by a t-test where $P<0.001$, which represents a statistically significant difference as between the mean of the normal patients as between the 25% and 75% interquartile range and the mean of the adenoma patients.

17. In terms of the different fold changes which are observed in the data in the specification and the subsequently submitted validation data, this is due to the fact that the data were generated from fundamentally different technologies which exhibit differences in their limits of detection. Depending on what the limit of detection is of a particular technology, this may result in differences in the actual fold change value which one obtains. The actual quantified levels of a molecule being measured cannot be compared directly as between two different types of technology. The data which appeared in the specification as originally filed were derived from gene specific RT-PCR, whereas the data which were submitted together with the Response filed in the '322 application on December 1, 2008 were generated from whole gene RNA microarrays. Although two entirely different technologies with different limits of detection were used, the ultimate outcome is consistent. That is, the mean level of expression of SEQ ID NO: 7 is increased in patients who have developed adenoma as opposed to patients who have not. While a variety of screening techniques are suitable for use in practicing the method of the '322 application, those skilled in the art would appreciate that the same technique should be used for both normal and test subjects to have a meaningful comparison and diagnosis.

18. I declare further that all statements made herein of our own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment; or both, under Section 1001 of Title 18 of the

United Stated Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Dated: 26 March 2010

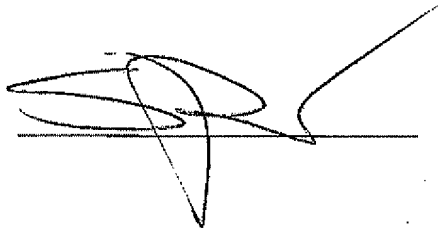


EXHIBIT 1

Dr. Susanne Pedersen

30 Michael Street North Ryde NSW Australia 2113

Citizenship: Denmark
Residency: Australia
Date of Birth: 4th January 1972

EDUCATION

1996-2000 Doctor of Philosophy (Molecular Biology)
Copenhagen Hospital, Denmark

1995-1996 Master of Sciences (Molecular Biology)
Copenhagen Hospital, Denmark

1991-1995 Bachelor of Sciences (Cell Biology)
Southern University of Denmark

POSTDOCTORAL EXPERIENCE

2003 - 2006 **Proteome Systems Limited, North Ryde NSW,**
SENIOR RESEARCH SCIENTIST – Key achievements

- Invented and patented novel enrichment techniques for isolation of immunogenic disease-specific proteins. These methods are now used across several Discovery projects to identify biomarkers of diagnostic and prognostic values.
- Responsible for design of experimental strategies for identification of pathogenic biomarkers for Cystic Fibrosis and Tuberculosis
- Current TB Discovery Project leader
- 2 years experience as a scientific adviser for a proteomic study of *Bacillus subtilis* for identification of new drug targets. This work includes project planning, budgeting, research evaluation and market analysis

Mar - Dec 2002 **PicoSep A/S, Odense, Denmark**
SCIENTIFIC ADVISER

- Scientific adviser for the commercialisation of novel surface chemistry technology
- Reporting directly to the CEO I provided experimental strategies to evaluate commercialisation opportunities for the technology
- Additional tasks involved setting up of a laboratory and training of staff

2000 - 2002 **Proteome Systems Limited, North Ryde NSW**
RESEARCH SCIENTIST

- Optimised sample preparation for 2DE profile analysis of ovarian cancer
- Optimised sample preparation and MS analysis for detection of low-abundant proteins
- Developed 2DE arraying methods for 'difficult' proteins, for e.g. membrane, acidic and alkaline proteins.

EDUCATION

1996-2000 **Clinical Department of Biochemistry, Copenhagen University Hospital (Rigshospitalet), Denmark.**

Dr. Susanne Pedersen

30 Michael Street North Ryde NSW Australia 2113

PhD Molecular Biology. Thesis Title: "*Translational control of insulin-like growth factor II mRNAs*".
Supervisor: Professor Finn Cilius Nielsen

1995-1995 **Clinical Department of Biochemistry, Copenhagen University Hospital (Rigshospitalet), Denmark.**
Master Degree, Molecular Biology. Project title: *Translational control of IGF-II mRNAs*
Supervisor: Professor Finn Cilius Nielsen

1991-1995 **Molecular and Biological Department, Southern University of Denmark, Odense, Denmark**
Bachelor Degree, Cell Biology. Project title: "Analysis of the *cea-kil* operon from the ColE1 plasmid". Supervisor: Dr. Kenn Gerdes

PATENTS

AU 2004/252182 A1	Method of isolating a protein
WO 2005/001480 A1	METHOD OF ISOLATING A PROTEIN
US 2007/0178541 A1	Method of isolating a protein
US 2009/0208535 A1	Novel Methods of Diagnosis of Treatment of P. Aeruginosa Infection and Reagents Therefor
AU 2005/256177 A1	Novel methods of diagnosis of treatment of P. aeruginosa infection and reagents therefor
WO 2006/000056 A1	NOVEL METHODS OF DIAGNOSIS OF TREATMENT OF P. AERUGINOSA INFECTION AND REAGENTS THEREFOR

PUBLICATIONS

- Pedersen SK, Sloane AJ, Prasad SS, Sebastian LT, Lindner RA, Hsu M, Robinson M, Bye PT, Weinberger R and Harry JL (2006). An immunoproteomic approach for identification of clinical biomarkers for monitoring disease: Application to cystic fibrosis. *Mol Cell Proteomics. 4 (8): 1052-1060*
- Sloane AJ, Lindner RA, Prasad SS, Sebastian LT, Pedersen SK, Robinson M, Bye PT, Nielson DW, Harry JL (2005). Proteomic analysis of sputum from adults and children with cystic fibrosis and from control subjects. *Am J Respir Care Med. 172 (11): 1416-1426*
- Hunt, S.M., Thomas, M.R., Sebastian, L.T., Pedersen, S.K., Harcourt, R.L., Sloane, A.J. and Wilkins, M.R. (2005). Optimal Replication of the Importance of Experimental Design for Gel-based Quantitative Proteomics. *Journal of. Proteome Research. 4 (3):809-819*
- Pedersen SK, Harry JL, Sebastian L, Baker J, Traini MD, McCarthy JT, Manoharan A., Wilkins MR, Gooley AA, Righetti PG, Packer NH, Williams KL and Herbert BR (2003). The unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins. *J Proteome Res.* 2 (3):303-11.
- Harcourt, R.L., Cole, R.A., Harry, E.J., Lindner, R.A., Pedersen, S.K., Prasad, S.S., Sebastian, L.T., Schulz, B.L., Sloane, A.J. and Harry, J.L. (2003). Proteomics: The

paradigm for biomarker and drug target discovery. *Journal of. Medical Technology*, **47** (11), Special Issue 'Clinical Protein Diagnostics: Direction to Proteomics'.

- Herbert, B.R., Pedersen, S.K., Harry, J.L., Sebastian, L., Traini, M.D., McCarthy, J.T., Wilkins, M.R., Gooley, A., Packer N.H., Williams, K., Righetti, G. and Grinyer, J. (2003). Mastering proteome complexity using two-dimensional gel electrophoresis. *PharmaGenomics* Sept 1: 22-36
- Herbert, B., Galvani, M., Hamdan, M., MacCarthy, J., Pedersen, S., Righetti, P.G (2001). Reduction and alkylation of proteins in preparation of two-dimensional map analysis: why, when and how? *Electrophoresis*, **22**: 2046-2057
- Herbert, B., Harry, J.L., Packer, N.H., Gooley, A.A., Pedersen, S.K., Williams, K (2001). What place for polyacrylamide in proteomics? TRENDS in Biotechnology, 19 (10): S3-S9
- Pedersen SK, Christiansen J, v. O. Hansen, T, Larsen MR, and Nielsen FC (2001). Human IGF-II leader 2 mediates internal initiation of translation. Biochem J. 363(Pt 1):37-44.
- Pedersen S, Celis JE, Nielsen J, Christiansen J and Nielsen FC (1996). Distinct Repression of translation by Wortmannin and Rapamycin. Eur. J. Biochem. 247:449-456.
- Breuner A, Jensen RB, Dam M, Pedersen S, Gerdes K. (1996). The centromere-like parC locus of plasmid R1. *Molecular Microbiology* **20** (3): 581-92.

GRANTS

Foundation for innovative new diagnostics (FIND), UNICEF/UNDP/WORLD Bank/ WHO, Special programme for research and training in tropical diseases (TDR):

Lodged 29 April 2005, Identification of in vivo expressed Mycobacterium tuberculosis proteins as biomarkers of infection. PI: Pedersen SK

LABORATORY SKILLS AND EXPERIENCE

- MALDI-TOF MS
- Post-source decay (PSD) MS
- MALDI TOF TOF
- R2 peptide purification
- MS analysis of phospho-proteins
- 2DE image analysis
- Immunoproteomics (immuno-capture and immuno-separation)
- Experienced in working with clinically biohazard samples
- Immuno-precipitation
- 2DE PAGE
- 1DE PAGE
- Western, Northern and Southern blotting
- Sample preparation for 2DE analysis, such as isolation of membrane proteins, acidic/alkaline proteins
- pI –based protein pre-fractionation techniques
- Coupling of various ligands (mRNA and proteins) to CnBr sepharose,
- Protein extraction for Gram positive bacteria
- Bioinformatics
- Protein modeling
- Guanidation of tryptic peptides
- Electrophoretic mobility shift assay
- DNA purification
- DNA sequencing
- PCR and DNA Cloning
- mRNA purification
- Expression and purification of His/GST tagged proteins
- Cell transfection
- Confocal laser microscopy
- Biotinylation of RNA
- UV crosslinking of DNA/RNA binding proteins

Dr. Susanne Pedersen

30 Michael Street North Ryde NSW Australia 2113

TALL POPPY CAMPAIGN

NOMINATION FORM 2006 Young Tall Poppy Science Awards

Outline of Nominee's Achievements – to be completed by the Nominator

As a co-founder and executive of Proteome Systems Ltd, an Australian biotechnology company, situated in Sydney and discovering disease diagnostics and drug targets, I would like to nominate one of our Project Leaders, Dr Susanne Pedersen, for the Young Tall Poppy Award.

Sanne joined us in 2000 at the end of her PhD in Denmark as she wanted to get her post doctoral experience in a company recognised for their proteomics technology expertise and who are committed to finding new disease diagnostics and making them accessible to the community. Her drive to experience new places and people in pursuit of her science in itself demonstrates her personality and independence.

In her first projects at the company, she pushed the boundaries of the technology to find low abundance biomarkers of ovarian cancer in blood and of lung infection of Cystic Fibrosis patients with Pseudomonas bacteria, the main cause of death in these patients.

This work resulted in patent applications for the methods she developed as well as for the diagnostic markers which were discovered. Many scientists fail to publish because of the patenting requirements of the commercial sector but Sanne's publication record is evidence of her persistence in following up with peer reviewed publications in recognised journals, after the patenting process.

Her obvious excellent science and focus quickly resulted in her appointment by the company to the position of Project Leader in which she ran two projects concurrently. One project was in collaboration with the University of Technology as part of an ARC Linkage grant in which she successfully supervised a post doctoral fellow and two PhD students in the proteomic analysis of bacterial cell division in a search for new antibiotic targets.

Simultaneously her breakthrough in using a novel immunological approach to finding low abundance proteins of the tuberculosis bacteria in blood and sputum formed the basis of the company embarking on developing a much needed antigen-based TB diagnostic test and has led to the funding of the project by the Bill Gates' Foundation for Innovative Diagnostics (FIND) and support of the Stop TB Partnership.

Sanne is an attractive, talented and intelligent young scientist who has made Australia her home and has taken out Permanent Residency here. Her potential as a role model for young women has been evidenced by her appearance in our company's promotional videos and media presentations. She has not yet had the opportunity in her current position to directly communicate her passion for science with the community and the award will hopefully give her the opportunity to do so.

Her key role as a dedicated and motivated innovator in the scientific team in our company I believe warrants her recognition by such an award as the Young Tall Poppy Science Award. It is often difficult within a company environment to access the type of rewards associated with such research in the academic sphere and I ask that she be considered for the Award in this light.

Dr. Susanne Pedersen

30 Michael Street North Ryde NSW Australia 2113

PICOSEP

PICOSEP A/S Forskerparken 10 5230 Odense M Telefon 63 15 29 20 Fax 63 15 32 15 www.picosep.com

To who it may concern

DATE
January 6th 2003

JOURNAL NO.

CONTACT
ARu

DIRECT PHONE
+45 5155 0660

E-MAIL
aru@picosep.com

Letter of recommendation

Susanne Karin Pedersen has been employed in PicoSep as Scientist from May 1st 2002 to November 30th 2002.

In this period Susanne has worked with fractionation of proteins in electrophoresis.

Susanne started out with implementation of a traditional 2D protein separation system. In a few month this system was implemented from scratch with all steps from sample preparation to image processing. The performance was well documented through a number of standard runs, also planed by Susanne.

However the real advantage for PicoSep has been Susanne's inventive and scientific approach to PicoSeps fractionation technology. Susanne has performed a lot of work on the design of both the fractionation system and on establishing and test protocols for the system. Susanne being co-inventor on a patent pending new protein fractionation system illustrates this inventive work.

Susanne is very keen in instructing the technical staff in the experimental test procedures and objectives. In addition Susanne has a charming and friendly personality, which helps establish a good working atmosphere in the laboratory. Documentation and reporting of obtained results is precise and reliable. I can only give Susanne my best recommendations.

Adam Rubin (CTO)
PicoSep A/S

Dr. Susanne Pedersen

30 Michael Street North Ryde NSW Australia 2113

Rigshospitalet
COPENHAGEN UNIVERSITY HOSPITAL
Dept. of Clinical Biochemistry 3013
Blegdamsvej 9
DK-2100 Copenhagen
Denmark

Phone: +45 35 45 3016/2223
Fax: +45 35 45 46 40

To Whom It May Concern.

Re.: FCN/cv

January 10, 2000

Letter of recommendation for - Susanne Kartin Pedersen
cpr: 04.01.72-1686

I shall hereby give Susanne my warmest recommendation.

Susanne came to the laboratory as graduate student and has since then worked with translation control and signal transduction. Moreover, she has recently focused on the identification of RNA binding proteins. So far her work has led to a single publication, but we expect that her Ph.D. thesis will provide basis for two additional papers on internal initiation of IGF-II mRNA and the function of *trans*-acting factors binding to IGF-II mRNA. Susanne is familiar with all the general recombinant DNA and RNA technology and basic protein chemistry, including mass-spectroscopy. Susanne is hard working, patient and easy to work with, and I am sure she will contribute, in a positive way, to the atmosphere in any laboratory. She works independently and conducts her experiments in a thorough and accurate manner.

If you need additional information please contact me.

Sincerely,

Finn Cilius Nielsen
Consultant, D.M. Sc.

EXHIBIT 2

Homo sapiens gene KIAA1199, encoding KIAA1199,

TABLE OF CONTENTS / OPEN CLOSE ALL PARAGRAPHS

SUMMARY

RefSeq annotates one representative transcript (NM included in AceView variant a), but Homo sapiens cDNA sequences in GenBank, filtered against clone rearrangements, coaligned on the genome and clustered in a minimal non-redundant way by the manually supervised AceView program, support at least 4 spliced variants.

AceView summary

Expression: According to AceView, this gene is expressed at high level, 2.1 times the average gene in this release. The sequence of this gene is defined by 243 GenBank accessions from 231 cDNA clones, some from bone (seen 39 times), hip (36), brain (16), colon (16), lung (15), nervous tumor (13), stomach (11) and 65 other tissues. We annotate structural defects or features in 6 cDNA clones.
Alternative mRNA variants and regulation: The gene contains 29 different gt-ag introns. Transcription produces 10 different mRNAs, 4 alternatively spliced variants and 6 unspliced forms. There are 2 non overlapping alternative last exons and 3 validated alternative polyadenylation sites (see the diagram). The mRNAs appear to differ by truncation of the 5' end, truncation of the 3' end, presence or absence of a cassette exon, overlapping exons with different boundaries. 102 bp of this gene are antisense to spliced gene MESDC2.aApr07, raising the possibility of regulated alternate expression.
Protein coding potential: 4 spliced and 2 unspliced mRNAs putatively encode good proteins, altogether 6 different isoforms (2 complete, 2 COOH complete, 2 partial), some containing a vacuolar domain [Psort2]; 1 of the 2 complete proteins appears to be secreted. The remaining 4 mRNA variants (4 unspliced; 1 partial) appear not to encode good proteins.
Function: There are 4 articles specifically referring to this gene in PubMed. Functionally, the gene has been tested for association to diseases (Hearing Loss; Kidney Neoplasms) and proposed to participate in a process (sensory perception of sound). Proteins are expected to localize in extracellular. A putative protein interactor has been described (PLXNA2).

Please quote: AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biology 2006, 7(Suppl 1):S12

Map on chromosome 15, links to other databases and other names ?

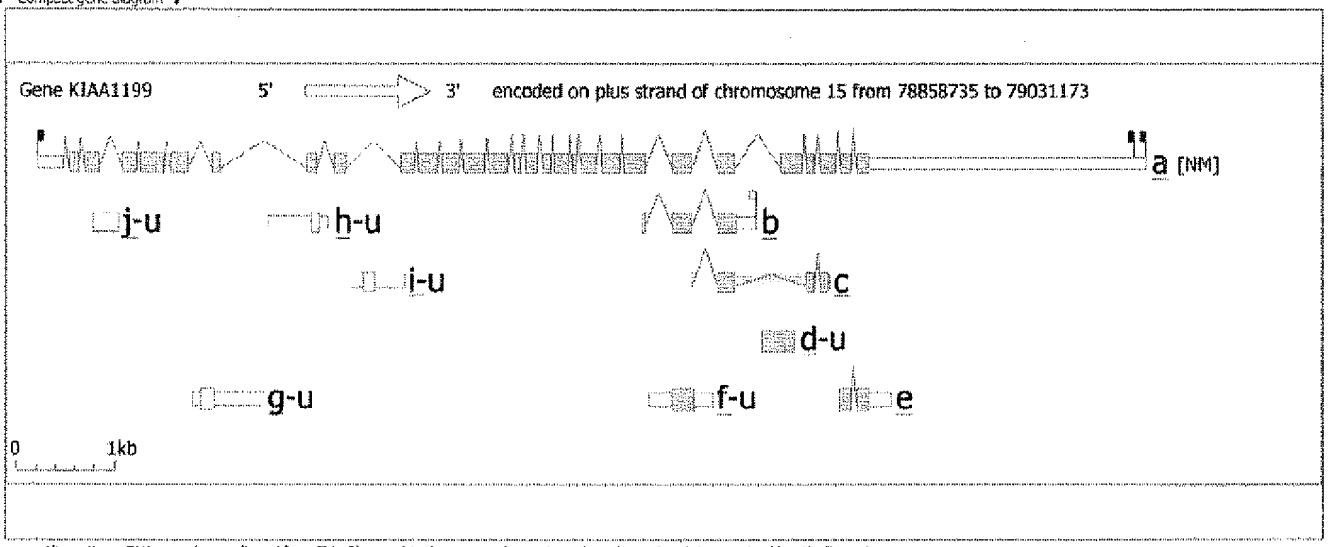
Map: This gene KIAA1199 maps on chromosome 15, at 15q24 according to Entrez Gene. In AceView, it covers 172.44 kb, from 78858735 to 79031173 (NCBI 36, March 2006), on the direct strand.
Links to: manual annotations from OMIM_608366, the SNP view, gene overviews from Entrez Gene 57214, GeneCards, expression data from ECgene, UniGene, molecular annotations from UCSC, or our GOLD analysis.
The previous AceView annotation is here.
Other names: The gene is also known as KIAA1199 or TMEM2L, LOC57214. It has been described as KIAA1199, transmembrane protein 2-like.

Closest AceView homologs in other species ?

The closest mouse gene, according to BlastP, is the AceView gene 9930013L23Rik (e=4 10^-38,).
The closest C.elegans genes, according to BlastP, are the AceView/WormGenes XD426, 4D357, 4D18, which may contain interesting functional annotation.
The closest A.thaliana gene, according to BlastP, is the AceView gene AT1G60500 (e=0.27,)

Complete gene on genome diagram: (in true scale, with colored introns)

Compact gene diagram

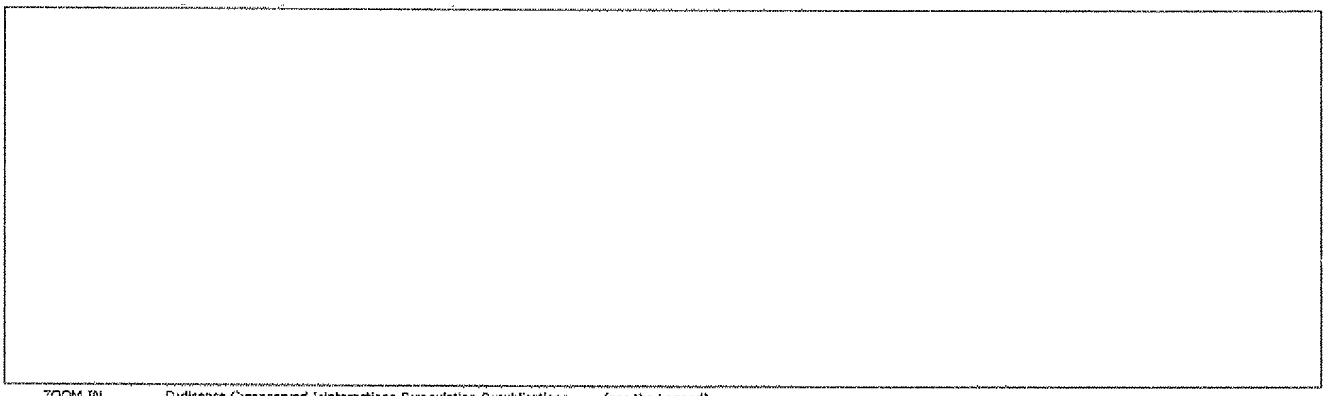


Alternative mRNAs are shown aligned from 5' to 3' on a virtual genome where introns have been shrunk to a minimal length. Exon size is proportional to length, intron height reflects the number of cDNA clones supporting each intron.
Mouse over the ending of each transcript gives tissues from which the supporting cDNAs were extracted. Click on any transcript to open the specific mRNA page, to see the exact cDNA clone support and eventual SNPs and to get details on tissues, sequences, mRNA and protein annotations. Details on tissue of origin for each intron and exon is available from the intron and exons table. Good predicted proteins are in pink or blue, yellow proteins may be partial or unconvincing. Proteins supported by a single continuous GenBank accession lead to underlining the name/ending of the variant. Names not underlined result from cDNA concatenation in the coding region and should be experimentally checked.

Caption

Sequences

What is known about the gene and its neighbors on chromosome 15q24



ZOOM IN D:disease,C:conserved,I:interactions,R:regulation,P:publications (see the Legend)

- Annotated mRNA diagrams
- Bibliography: 6 articles in PubMed

To mine knowledge about the gene, please click the 'Gene Summary' or the 'Function and related genes' tab at the top of the page. The 'Gene Summary' page includes all we learnt about the gene, functional annotations of neighboring genes, maps, links to other sites and the bibliography. The 'Function and related genes' page includes Diseases (D), Pathways, GO annotations, conserved domains (C), interactions (I) reference into function, and pointers to all genes with the same functional annotation.
To compare all variants, their summarized annotations, introns and exons, or to access any sequence, click the 'Alternative mRNAs features' tab. To see a specific mRNA variant diagram, sequence and annotation, click the variant name in the 'mRNA' tab. To examine expression data from all cDNAs clustered in this gene by AceView, click the 'Expression tissue'.

If you know more about this gene, or found errors, please share your knowledge. Merci ! 0

EXHIBIT 3

FIGURE 1

TRANSCRIPTIONAL ACTIVITY ACROSS THE KIAA1199 GENE LOCUS

Expression levels of NCBI KIAA1199 RefSeq was measured using Affymetrix HuGene ST 1.0 Genechip microarrays by monitoring RNA hybridization levels to 25 oligonucleogide probes complementary to regions in the KIAA1199 gene by processing the Genechips according to manufacturer's instructions for the Affymetrix HuGene ST 1.0 array using RNA extracted from colon tissue specimens from 30 normals (left bars), 21 adenomas (middle bars) and 21 colorectal cancers (right bars). Further information regarding methodology is available in Figure 3.

FIGURE 1

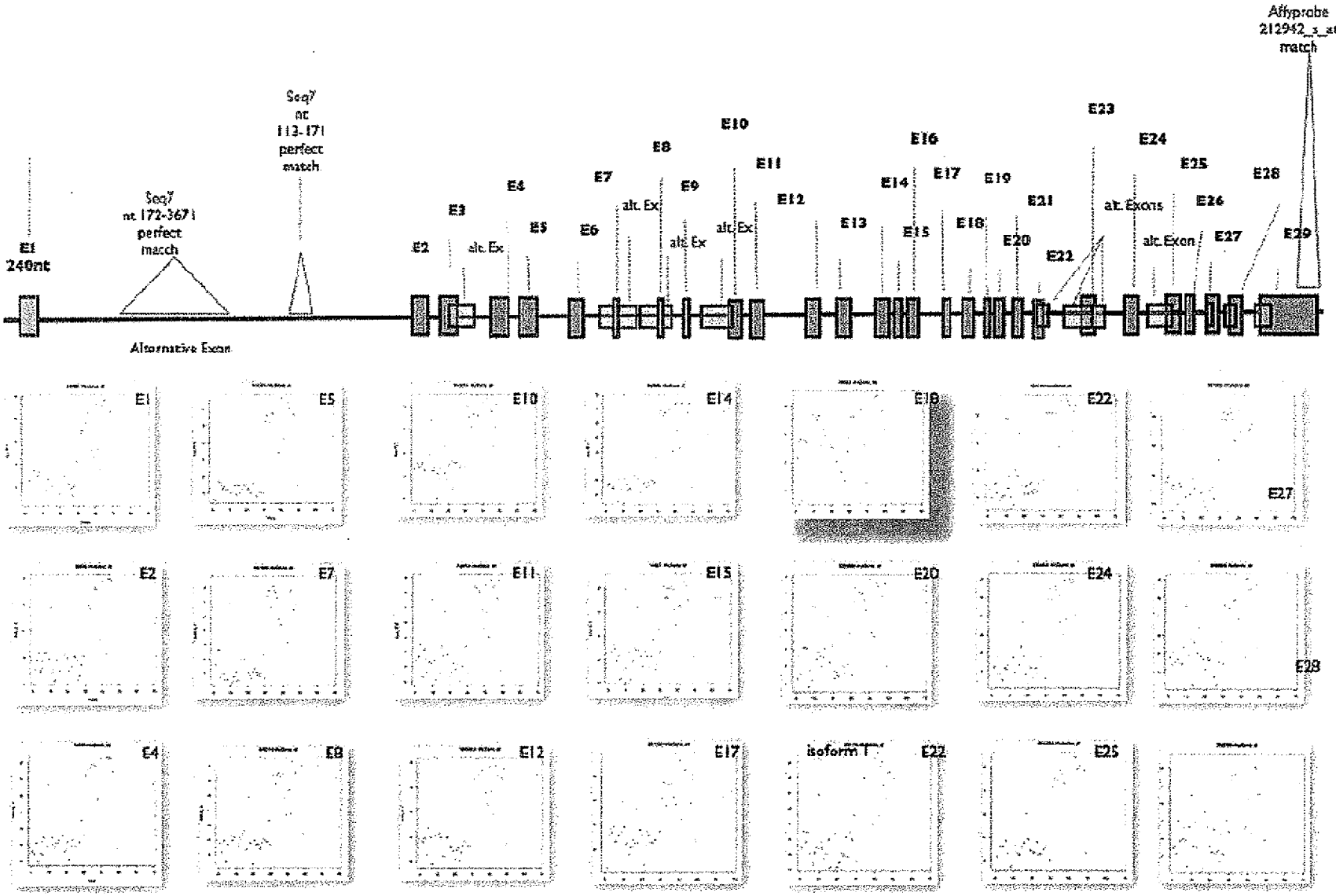
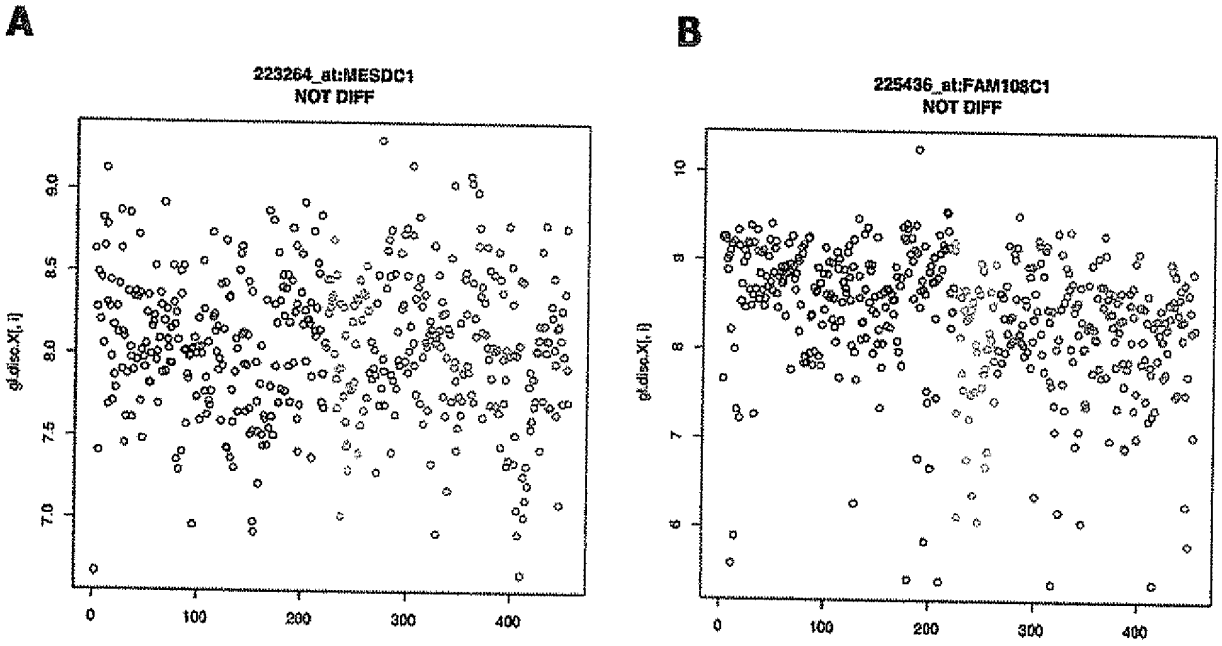

Affyprobe
212942_s_at
match
Seq7
nt 113-171
perfect
match
Seq7
nt 172-3671
perfect
match
E1
240nt
Alternative Exon
E2
alt. Ex
E3
E4
E5
E6
E7
alt. Ex
E8
E9
alt. Ex
E10
alt. Ex
E11
E12
E13
E14
E15
E16
E17
E18
E19
E20
E21
E22
E23
alt. Exons
E24
alt. Exon
E25
E26
E27
E28
E29
E1
E5
E10
E14
E18
E22
E27
E2
E7
E11
E15
E20
E24
E28
E4
E8
E12
E17
isoform 1
E22
E25

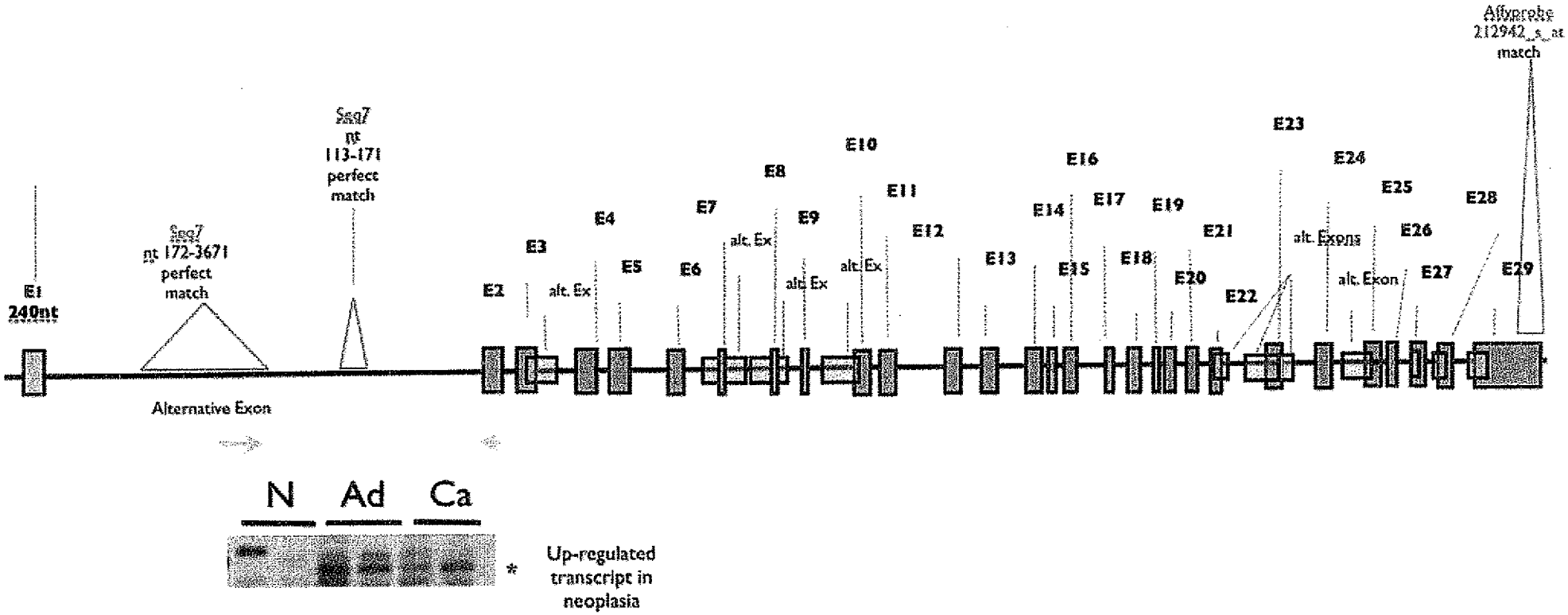
FIGURE 2



TRANSCRIPTIONAL ACTIVITY OF GENES FLANKING KIAA1199 GENE LOCUS

(A) Gene activity of MESDC1 (downstream neighbor gene to KIAA1199 gene locus - located 50,993nt downstream) and (B) FAM108C1 (upstream neighbor gene to KIAA1199 gene locus – located 23,722nt upstream) were determined by measuring level of RNA hybridization to probesets 223264_at and 225436_a, respectively, using Affymetrix' U133 Plus 2.0 Genechips and RNA extracted from colon tissue specimens from 222 normals (black), 42 subjects with IBD (green), 29 subjects with adenomas (blue) and 161 subjects with colorectal cancer (red).

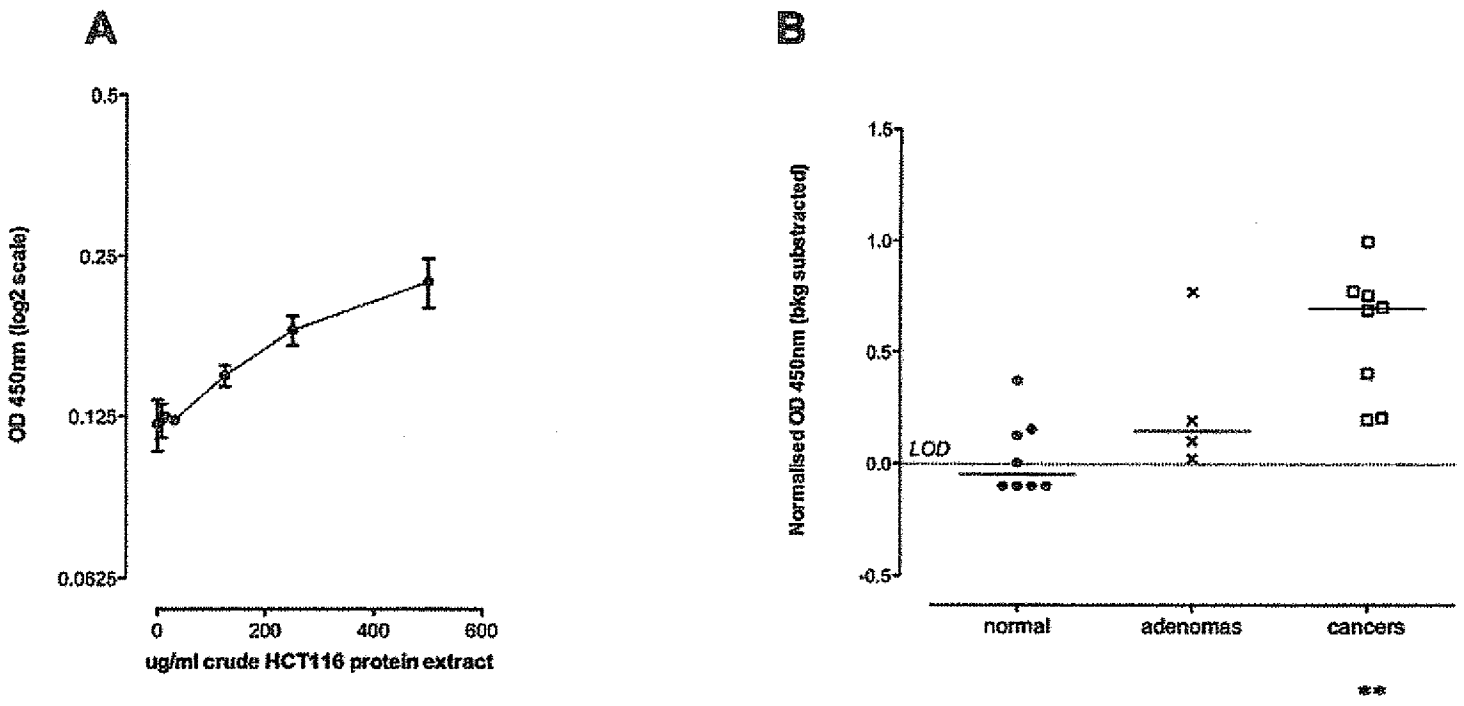
FIGURE 3



SEQID_7 IS A TRANSCRIPTIONAL DERIVATIVE FROM THE KIAA1199 GENE LOCUS

To demonstrate transcriptional relationship between SEQID_7 and the NCBI RefSeq KIAA1199 gene locus, a PCR assay using a forward primer residing in the 3'end of SEQID_7 (right pointing arrow) and a reverse primer residing in the 5'end of NCBI KIAA1199 RefSeq Exon 2. The PCR assays were used to screen RNA extracted from colon tissue specimens from 2 normal, 2 adenomas and 2 cancers. A dominant band was observed in neoplastic tissues. DNA sequencing confirmed the PCR product to represent both SEQID_7 and KIAA1199 NCBI RefSeq Exon 2 sequences.

EXHIBIT 4



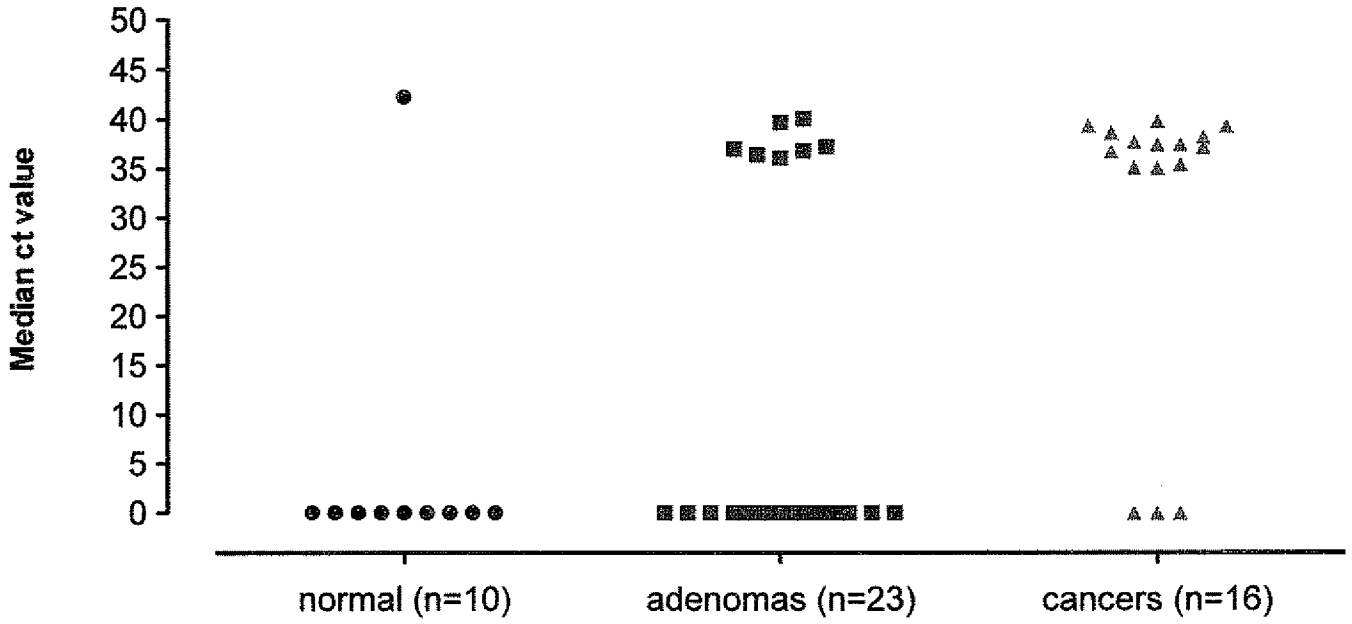
DETECTION OF KIAA1199 PROTEIN IN STOOL SPECIMENS

An indirect ELISA using a monoclonal anti-KIAA1199 antibody was used to develop an immunoassay for detection of KIAA1199 protein. (A) 2-fold serial dilution of HCT116 cell line crude protein. (B) Stool protein extracted from 7 normal subjects (black circles), 4 adenomas (crisscross) and 8 colorectal cancer subjects (squares) were analysed as duplicates by Indirect ELISA using 10µg/ml monoclonal anti-KIAA199 antibody (Abnova) and 1:5,000 diluted HRP-conjugated sheep anti-mouse IgG antibody. Immunocomplexes were detected using TMB as recommended by manufacturer followed by OD-450nm readings. Signals are displayed as background subtracted mean values. **; Mann-Whitney non-parametric t-test: p = 0.002 (N.S.).

Part 2

KIAA1199 E23/E24 TaqMan Plasma

(KIAA1199 transcript detected in 1/10 normals, 7/23 adenomas and 13 out of 16 cancers)



RNA was extracted from 2mL of plasma using the Qiagen kit "Isolation of circulating nucleic acids from plasma/sera" as recommended. The resulting RNA was converted to cDNA as standard procedure and the level of endogenous KIAA1199 RNA transcript was measured using a commercial KIAA1199 TaqMan qPCR kit (Applied Biosystems' cat no: Hs01552116_m1) as recommended. Samples were analysed as triplicates and medium ct-values were calculated only on samples in which a signal was detected in all three replicates.

Diagnostic mRNA Expression Patterns of Inflamed, Benign, and Malignant Colorectal Biopsy Specimen and their Correlation with Peripheral Blood Results

Orsolya Galamb,[1,3] Ferenc Sipos,[1] Norbert Solymosi,[1] Sándor Spisák,[1] Tibor Krenács,[2] Kinga Tóth,[1] Zsolt Tulassay,[1,3] and Béla Molnár[1,3]

[2]2nd Department of Medicine and [1]1st Department of Pathology and Experimental Cancer Research, Semmelweis University; and [3]Molecular Medicine Research Unit, Hungarian Academy of Sciences, Budapest, Hungary

Abstract

Purpose: Gene expression profile (GEP)–based classification of colonic diseases is a new method for diagnostic purposes. Our aim was to develop diagnostic mRNA expression patterns that may establish the basis of a new molecular biological diagnostic method.
Experimental Design: Total RNA was extracted, amplified, and biotinylated from frozen colonic biopsies of patients with colorectal cancer (n = 22), adenoma (n = 20), hyperplastic polyp (n = 11), inflammatory bowel disease (n = 21), and healthy normal controls (n = 11), as well as peripheral blood samples of 19 colorectal cancer and 11 healthy patients. Genome-wide gene expression profile was evaluated by HGU133plus2 microarrays. To identify the differentially expressed features, the significance analysis of microarrays and, for classification, the prediction analysis of microarrays were used. Expression patterns were validated by real-time PCR. Tissue microarray immunohistochemistries were done on tissue samples of 121 patients.
Results: Adenoma samples could be distinguished from hyperplastic polyps by the expression levels of nine genes including ATP-binding cassette family A, member 8, insulin-like growth factor 1 and glucagon (sensitivity, 100%; specificity, 90.91%). Between low-grade and high-grade dysplastic adenomas, 65 classifier probesets such as aquaporin 1, CXCL10, and APOD (90.91/100) were identified; between colorectal cancer and adenoma, 61 classifier probesets including axin 2, von Willebrand factor, tensin 1, and gremlin 1 (90.91/100) were identified. Early- and advanced-stage colorectal carcinomas could be distinguished using 34 discriminatory transcripts (100/66.67).
Conclusions: Whole genomic microarray analysis using routine biopsy samples is suitable for the identification of discriminative signatures for differential diagnostic purposes. Our results may be the basis for new GEP-based diagnostic methods. (Cancer Epidemiol Biomarkers Prev 2008;17(10):2835–45)

Introduction

Colorectal cancer is one of the most frequent cancers in the world with very high mortality. According to WHO data, ~945,000 new colorectal cancer cases are registered worldwide, and almost 492,000 colorectal cancer–related deaths occur every year (1). Hence, the early diagnosis, the discrimination between genetically and expressionally different tumors, and in view of these, the enhancement of therapies, become necessary. The 5-year survival data also emphasize the importance of an early diagnosis of colorectal cancer. The 5-year survival rate is 80% to 90% in early colorectal cancer, 60% in case of nodal involvement, and <10% in metastatic colorectal cancer.

According to the widely accepted adenoma-dysplasia-carcinoma sequence, most of the colorectal cancer develop on the basis of villous adenomas (2, 3). Recently published, however, was the concept of a "serrated neoplasia pathway" referring to a pattern of progression of colorectal cancer that involves hyperplastic polyps and serrated adenomas (4). The serrated pathway culminates in colorectal cancers with DNA microsatellite instability, mutation of BRAF, and extensive DNA methylation (5-7). Iino et al. (8) suggested that MSI-L hyperplastic polyps may be precursors of the subset (10%) of colorectal cancers showing the MSI-L phenotype.

Gene expression analysis of colon biopsies using high-density oligonucleotide microarrays may help to detect such gene expression patterns that would establish the basis for new molecular biological diagnostic methods. Utilization of mRNA expression microarray data for diagnostic purposes has already begun. More and more scientific studies appear to focus on the gene expression background of colorectal cancer progression and metastasis development (9-18), characterization of colorectal cancer subtypes according to mRNA expression (12, 18, 19), the correlation of gene expression profile with clinicopathologic variables (12, 18, 20, 21), and mRNA expression–based prognosis (22). In addition to the surgical and biopsy tissue samples, mRNA expression analysis of peripheral blood samples may also play a crucial role in the establishment of early molecular-based diagnostics and prognostics of tumorous diseases (23-27). The handling and the evaluation of the huge amount of

Received 3/14/08; revised 6/11/08; accepted 7/28/08.
Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).
Requests for reprints: Orsolya Galamb, 2nd Department of Medicine, Semmelweis University, Szentkirályi str. 46, 1088 Budapest, Hungary. Phone: 36-1-266-00-26; Fax: 36-1-266-08-16. E-mail: orsgt@yahoo.com

doi:10.1158/1055-9965.EPI-08-0231

Table 1. Number of patients in the different disease groups

Group	Biopsy samples, n = 85, original set			Biopsy samples, n = 92, independent set	Blood samples, n = 31
	Affymetrix microarray	Taqman RT-PCR	Tissue microarray	Tissue microarray	Affymetrix microarray
Adenoma with low-grade dysplasia	9	6	—	—	—
Adenoma with high-grade dysplasia	11	6	—	—	—
CRC Dukes A-B	10	6	2	20	7
CRC Dukes C-D	12	4	2	21	12
Normal	11	5	9	21	11
Hyperplastic polyp	11	—	—	—	—
Ulcerative colitis	12	7	10	8	—
Crohn's disease	9	—	6	16	—
Undeterminate IBD	—	—	—	7	—
Total patient numbers	85	34	29	93	30

Abbreviation: CRC, colorectal cancer.

data collected by microarray analyses require an extensive bioinformatical background. Multivariate statistical analysis is needed for the development of automatic diagnostic disease classification methods.

We have previously reported the discriminative mRNA expression signatures between colorectal cancer versus normal, adenoma versus normal, inflammatory bowel disease (IBD) versus normal samples, and between the early and advanced stages of colorectal cancer (28). However, the gene expression profile–based classification of colonic diseases for diagnostic purposes has not yet been solved. The results of the HGU133 Plus 2.0 whole genomic microarrays—which were also used in our study—in colorectal diseases have been published by only five research groups (29-33), and only two of them used biopsy samples (29, 30). Using Affymetrix microarrays, high-throughput disease-specific marker screening can be done. Our aims in this study were to develop diagnostic mRNA expression patterns for the objective classification of inflammatory, benign, and malignant colorectal diseases, and to compare the gene expression background of adenomas and hyperplastic polyps as the possible points of origin of colorectal cancer. Furthermore, we analyzed the presence of certain local colorectal cancer markers in peripheral blood that had been identified while using biopsy samples. This is necessary for the development of blood-based, disease-specific diagnostic screening.

Materials and Methods

Patients and Samples. After the informed consent of untreated patients, colon biopsy samples were taken during endoscopic intervention and stored in RNALater Reagent (Qiagen, Inc.) at • 80jC. Additionally, 9 mL of peripheral blood samples of untreated patients were taken into Paxgene Blood RNA Tubes (Qiagen) before colonoscopy. The blood samples were also stored at • 80jC. Altogether, 377 tissue samples (85 fresh frozen and 292 formalin-fixed paraffin-embedded tissue samples) and peripheral blood samples of 19 colorectal cancer and 11 healthy patients were analyzed in our study, as well as the blood smears of 10 healthy and 10 colorectal cancer patients. Total RNA was extracted, and Affymetrix microarray analysis was done on the biopsies of patients with tubulovillous/villous adenomas (n = 20, 11 with high-grade dysplastic and 9 with low-grade dysplasia), colorectal adenocarcinoma (n = 22), hyperplastic polyps (n = 11), and healthy normal controls (n = 11), as well as from peripheral blood samples of 19 patients with colorectal cancer and 11 healthy patients. Fifty-two microarrays (8 normal, 15 adenoma, 15 colorectal cancer, 14 IBD) had been hybridized earlier; their data files were used in a previously published study using different comparisons (28) and are available in the Gene Expression Omnibus database (series accession number: GSE4183). The data sets of the newly hybridized 63 microarrays are registered in the GSE10714 (33 microarrays from biopsy samples: 3 normal, 11 hyperplastic polyps, 5 adenoma, 7 colorectal cancer, 7 IBD) and in the GSE10715 (30 microarrays from blood samples: 19 colorectal cancer and 11 normal) serial accession numbers. The diagnostic groups and the number of patients in each group are represented in Table 1. Detailed patient specification is described in Supplementary Table S1.

Methods

mRNA Expression Microarray Analysis. Total RNA was extracted using the RNeasy Mini Kit (Qiagen) for biopsy samples and the Paxgene Blood RNA Kit (Qiagen) for peripheral blood samples according to the manufacturers' instructions. The isolated peripheral blood RNA samples were concentrated using the GeneChip Blood RNA Concentration Kit (Affymetrix, Inc.). The quantity and the quality of the isolated RNA were tested by measuring the absorbance and agarose gelelectrophoresis or capillary gelelectrophoresis using the 2100Bioanalyzer and RNA 6000 Pico Kit (Agilent, Inc.). Biotinylated cRNA probes were synthesized from 5 to 8 Ag total RNA and fragmented using the One-Cycle Target Labeling and Control Kit[4] according to the Affymetrix description. In case of peripheral blood RNA samples, 5 Ag total RNA was used for cRNA probe synthesis, and during reverse transcription Globin Reduction PNA oligomers (Applied Biosystems) were applied to reduce the amount of globin transcripts. Ten micrograms of each fragmented cRNA sample were hybridized into HGU133 Plus2.0 array

[4] https://www.affymetrix.com/support/downloads/manuals/expression_s2_manual.pdf

Table 2. Discriminatory PAM transcripts between the diagnostic groups

Group vs group	Minimum no. of discriminatory transcripts	Sensitivity (%)	Specificity (%)	Including transcripts:
Adenoma vs hyperplastic polyp	9	100	90.91	ABCA8, KIAA1199, GCG, MAMDC2, C2orf32, 229670_at, IGF1, PCDH7, PRDX6
IBD vs normal	3	100	100	REG1A, MMP3, CHI3L1
Adenoma vs CRC	61	90.91	100	GREM1*, DDR2*, GUCY1A3*, TNS1, ADAMTS1, FBLN1, FLJ38028, RDX, FAM129A, ASPN, FRMD6, MCC, RBMS1*, SNAI2, MEIS1, DOCK10, PLEKHC1, FAM126A, TBC1D9, VWF, DCN, ROBO1, MSRB3, LATS2, MEF2C*, IGFBP3*, GNB4, RCN3, AKAP12, RFTN1, 226834_at, COL5A1, GNG2, NR3C1*, SPARCL1, MAB21L2, AXIN2, 236894_at, AEBP1, AP1S2, C10orf56, LPHN2, AKT3, FRMD6, COL15A1, CRYAB, COL14A1, LOC286167, QKI, WWTR1, GNG11, PAPPA, ELDT1
IBD vs CRC	20	100	95.24	227458_at, INDO, CXCL9, CCR2, CD38, RARRES3, CXCL10, FAM26F*, TNIP3, NOS2A, CCRL1, TLR8, IL18BP, FCRL5, SAMD9L, ECGF1, TNFSF13B, GBP5, GBP1
CRC-B vs CRC-CD	34	100	66.67	TMEM37*, IL33, CA4*, CCDC58, CLIC6, VSNL1*, ESPN, APCDD1, C13orf18, CYP4X1, ATP2A3, LOC646627, MUPCDH, ANPEP, C1orf115, HSD3B2, GBA3, GABRB2, GYLTL1B, LYZ, SPC25, CDKN2B, FAM89A, MOGAT2, SEMA6D, 229376_at, TSPAN5, IL6R, SLC26A2
Adenoma with low-grade dysplasia vs adenoma with high-grade dysplasia	65	90.91	100	SI, DMBT1, CFI*, AQP1, APOD, TNFRSF17, CXCL10, CTSE, IGHA1, SLC9A3, SLC7A1, BATF2, SOCS1, DOCK2, NOS2A, HK2, CXCL2, IL15RA, POU2AF1, CLEC3B, ANI3BP, MGC13057, LCK*, C4BPA, HOXC6, GOLT1A, C2orf32, IL10RA, 240856_at, SOCS3, MEIS3P1, HIPK1, GLS, CPLX1, 236045_x_at, GALC, AMN, CCDC69, CCL28, CPA3, TRIB2, HMGA2, PLCL2, NR3C1, EIF5A, LARP4, RP5-1022P6.2, PHLDB2, FKBP1B, INDO, CLDN8, CNTN3, PBEF1, SLC16A9, CDC25B, TPSB2, PBEF1, ID4, GJB5, CHN2, LIMCH1, CXCL9, MFAP4
UC vs CD	58	77.78	100	CCNG2, SLC44A4, DDAH1, TOB1, 231152_at, MKNK1, CEACAM7*, 1562836_at, CDC42SE2, PSD3, 231169_at, IGL@*, GSN, GPM6B, CDV3*, PDPK1, ANP32E, ADAM9, CDH1, NLRP2, 215777_at, OSBPL1, VNN1, RABGAP1L, PHACTR2, ASH1L, 213710_s_at, ZNF3, FUT2, IGHA1, EDEM1, GPR171, 229713_at, LOC643187, FLVCR1, SNAP23*, ETNK1, LOC728411, POSTN, MUC12, HOXA5, SIGLEC1, LARP5, PIGR, SPTBN1, UFM1, C6orf62, WDR90, ALDH1A3, F2RL1, IGHV1-69, DUOX2, RAB5A, CP
Hyperplastic polyp vs normal	15	100	100	SLC6A14, ARHGEF10, ALS2, IL1RN, SPRY4, PTGER3, TRIM29, SERPINB5, 1560327_at, ZAK, BAG4, TRIB3, TTL, FOXQ1, UGT2A3
Adenoma with low-grade dysplasia vs normal	3	100	100	KLK11, KIAA1199, FOXQ1
Adenoma with high-grade dysplasia vs normal	3	100	100	CLDN8, ABCA8, PYY
Adenoma vs normal	3	100	100	KIAA1199, FOXQ1, CA7
CRC vs normal	5	100	100	VWF, IL8, CHI3L1, S100A8, GREM1

Abbreviations: CD, Crohn's disease; UC, ulcerative colitis.
*Transcripts represented by more probesets.

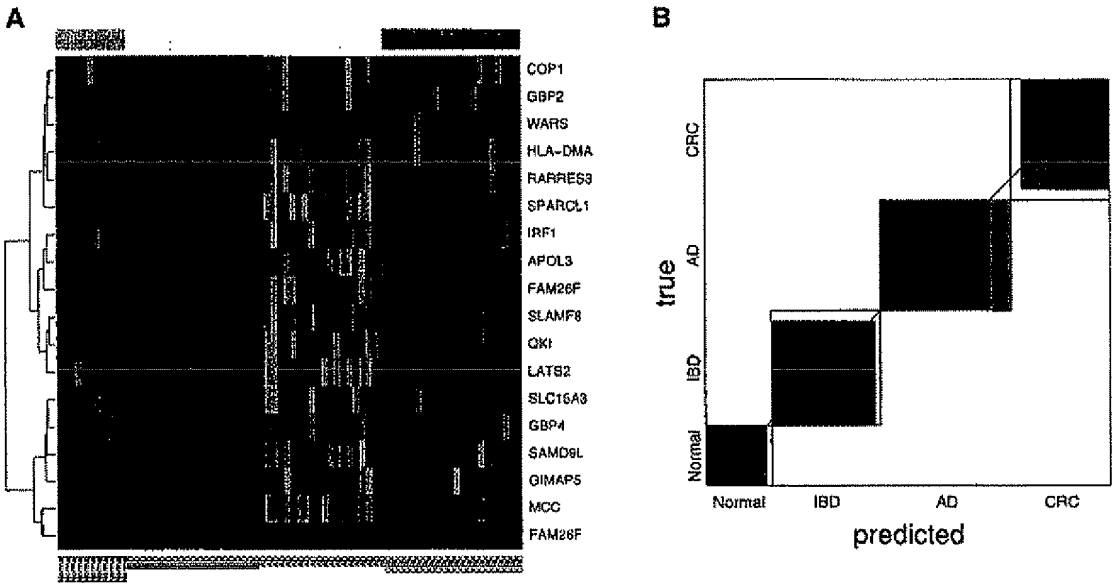


Figure 1. Random forest. A. Heat map of the diagnostic groups separated using the random forest classification method. The heatmap visualizes the expression level of genes (rows) that were selected as classifiers using the random forest supervised machine learning method. One can realize the difference of gene expression according to the different diagnostic groups (columns). B. Agreement plot for visualization of the confusion matrix of the true and the predicted classes. The agreement plot is the representation of the strength of agreement in the confusion matrix of the observed (true) and predicted classes. The prediction of each sample was based on the classifier using the genes presented on heatmap. Black areas show the observed agreement positioned within larger rectangles representing the maximum possible agreement, given the marginal totals. Gray areas represent the degree of disagreement. AD, adenoma; CRC, colorectal cancer; IBD, inflammatory bowel disease.

(Affymetrix) at 45jC for 16 h. The slides were washed and stained using Fluidics Station 450 and an antibody amplification staining method according to the manufacturer's instructions. The fluorescent signals were detected by a GeneChip Scanner 3000.

Statistical Evaluation of mRNA Expression Profiles

Preprocessing and Quality Control. Quality control analyses were done according to the suggestions of The Tumour Analysis Best Practices Working Group (34). Scanned images were inspected for artifacts, and the percentage of present calls (>25%) and control of the RNA degradation were evaluated. Based on the evaluation criteria, all biopsy measurements fulfilled the minimal quality requirements. The Affymetrix expression arrays were preprocessed by gcRMA with quantile normalization and median polish summarization. The data sets are available in the Gene Expression Omnibus databank for further analysis,[5] series accession numbers GSE4183, GSE10714, and GSE10715.

Further Analyses. To identify differentially expressed features, significance analysis of microarrays was used. The nearest shrunken centroid method (prediction analysis of microarrays) was applied for sample classification from gene expression data. For gene selection, the random forest classification algorithm was used (35), whereas the .632+ bootstrap method was applied to estimate the prediction error rate (36). The confusion matrix of the true and the predicted classes was visualized on agreement plots (37). The preprocessing, data mining, and statistical steps were done using R-environment with Bioconductor libraries.

Taqman Real-time PCR. TaqMan real-time PCR (RT-PCR) was used to measure the expression of 26 selected genes using an Applied Biosystems Micro Fluidic Card System. The selected genes belonged to the prediction analysis of microarrays top 200 genes in the colorectal cancer versus normal, adenoma versus normal, and IBD versus normal comparisons, and validated Taqman assays were available. The measurements were done using an ABI PRISM 7900HT Sequence Detection System as described in the product's user guide.[6] The data analysis was described earlier (28). For data analysis, the SDS 2.2 software was used.

Tissue Microarray Analysis and Blood Smear Immunocytochemistry. Cores of 1-mm diameter were collected from selected areas of formalin-fixed, paraffin-embedded tissue blocks made from 89 early colorectal cancer (stage Dukes B), 57 advanced colorectal cancer (stage Dukes C and D), 84 IBD (32 Crohn's disease, 40 ulcerative colitis, and 12 undeterminate IBD), and 62 normal colon samples

[5] http://www.ncbi.nlm.nih.gov/geo/

[6] http://www.appliedbiosystems.com

Table 3. Correlation between colorectal cancer versus normal biopsy and peripheral blood results

Gene symbol	Probeset ID
Up-regulated in CRC compared with normal in both biopsy and blood samples	
TPM4	212481_s_at
SESTD1	226763_at
TTYH3	224674_at
TIMP1	201666_at
CD44	212014_x_at
TM9SF4	212194_s_at
PIM3	224739_at
PELO	218472_s_at
C6orf145	212923_s_at
SFXN3	217226_s_at
MYL9	201058_s_at
CD44	210916_s_at
CD44	204490_s_at
VCAN	221731_x_at
CD44	209835_x_at
VCAN	204620_s_at
VCAN	211571_s_at
TGFBI	201506_at
PLXND1	38671_at
TKT	208700_s_at
VCAN	215646_s_at
PF4	206390_x_at
CD44	1557905_s_at
IFITM3	212203_x_at
S100A11	200660_at
NA	228910_at
G6PD	202275_at
AP1M1	223025_s_at
ZC3H12A	218810_at
FSCN1	210933_s_at
NDE1	218414_s_at
IER3	201631_s_at
PEA15	200787_s_at
PTP4A3	206574_s_at
IMPDH1	204169_at
PRKCDBP	213010_at
DDEF1	224796_at
ESAM	225369_at
CCDC85B	204610_s_at
MGC7036	227983_at
IFITM2	201315_x_at
IFITM1	201601_x_at
COL18A1	209082_s_at
RAB31	217762_s_at
FLNA	214752_x_at
TMEM158	213338_at
CTSK	202450_s_at
ENC1	201340_s_at
ICAM1	202638_s_at
INTS1	212212_s_at
PI3	203691_at
NA	227041_at
Down-regulated in CRC compared with normal in both biopsy and blood samples	
SLC26A2	224959_at
NA	227682_at
UGDH	203343_at
Up-regulated in CRC compared with normal in biopsy, down-regulated in blood samples	
RANBP2	201712_s_at
DNAJC10	225174_at
CRKRS	225694_at
SLC39A6	202088_at
SLC39A6	202089_s_at
DIS3	222607_s_at
ELK3	221773_at
DNAJC10	229588_at

(Continued on the following page)

Table 3. Correlation between colorectal cancer versus normal biopsy and peripheral blood results (Cont'd)

Gene symbol	Probeset ID
RANBP5	211953_s_at
IL8	202859_x_at
RANBP2	226922_at
SACS	213262_at
DNAJC10	221782_at
POT1	204354_at
GALNACT-2	218871_x_at
HS2ST1	203284_s_at
XPOT	212160_at
Down-regulated in CRC compared with normal in biopsy, up-regulated in blood samples	
MTMR11	205076_s_at
ETHE1	204034_at
SULT1A3	209607_x_at
C9orf19	225604_s_at
AGXT2L2	226519_s_at
SULT1A2	207122_x_at
FCGRT	218831_s_at
TRPM6	240389_at
SULT1A2	211385_x_at
SULT1A1	203615_x_at
ACADVL	200710_at
C22orf16	224932_at

of 122 patients and placed into recipient blocks. Tissue sections of 5-Am thickness were cut from the blocks and immunostained using the following antibodies: rabbit anti-human osteopontin (1:2,000 dilution; Chemicon), anti-osteonectin antibody (1:1,000 dilution; Chemicon), rabbit antihuman biglycan (1:200 dilution; Atlas), mouse anti-human collagen type IVa1 (1:300 dilution; Abcam, clone: COL-94), mouse anti-human vascular endothelial growth factor (1:2,000 dilution; Zymed, clone: VG 1), mouse anti-human von Willebrand factor (1:20; Dako, clone: F8/86), and mouse anti-human platelet-endothelial cell adhesion molecule 1 (1:40; Dako, clone: JC70A). Signal conversion was achieved using the EnVision+ kit (Dako) followed by 3,3}-diaminobenzidine–hydrogen peroxidase chromogen-substrate kit (Dako). Immunostained tissue microarray (TMA) slides were digitalized using a high-resolution Mirax Desk instrument (Zeiss) and analyzed with the Mirax TMA Module software (Zeiss). Protein expression was evaluated using an empirical scale considering intensity and occupied subcellular compartments of epithelial/carcinoma cells or lamina propria cells. For statistical analysis, Pearson's m^2 test and Fisher's exact test were done.

Blood smears of 10 healthy and 10 colorectal cancer patients were also immunostained using an anti-osteonectin antibody (1:1,000 dilution; Chemicon) and Alexa Fluor 488 F(ab|)2 fragment of goat anti-mouse IgG. The total and osteonectin-positive cells in 50 fields of view with 30• magnification were counted in each sample. For statistical analysis, a t test was done to evaluate the difference of osteonectin-positive/total cell number ratios between colorectal cancer and normal blood smears.

Results

Classifiers between the Main Diagnostic Groups. The minimal number of discriminatory transcripts with high specificity and sensitivity values was determined

using prediction analysis of microarrays in each comparison. Adenoma samples were distinguished from hyperplastic polyps by 100% sensitivity and 90.91% specificity, according to the expression level of minimally nine genes including ATP-binding cassette family A, member 8, insulin-like growth factor 1 and glucagon. Sixty-one classifier probesets were identified between colorectal cancer and adenoma, including axin 2, von Willebrand factor, tensin 1, and gremlin 1 (sensitivity, 90.91% and specificity, 100%). IBD and normal biopsies could be distinguished by 100% sensitivity and specificity using only three classifiers (REG1A, MMP3, and CHI3L1). According to the expression of 20 transcripts (such as INDO, CXCL9, CCR2, CD38, RARRES3, and CXCL10 transcripts), IBD and colorectal cancer samples could be separated by 100% sensitivity and by 95.24% specificity. Further details can be seen in Table 2.

Beside pair-wise comparisons, the random forest classification was also done to distinguish between the above-mentioned diagnostic groups (Fig. 1). The estimated prediction error was 12.9%. The main diagnostic groups could be distinguished according to the mRNA expression levels of 18 genes, including cell cycle and cell proliferation regulatory genes (retinoic acid responder 3, LATS large tumor suppressor homologue 2, mutated in colorectal cancers, WARS), COP1 apoptosis gene, HLA-DMA, APOL3, GBP2, SLAMF8 inflammatory response related genes, SPARC-like 1 calcium-binding extracellular matrix gene, SLC15A3 oligopeptide transporter, as well as IFN regulatory factor 1, and quaking homologue transcription and mRNA processing genes. The exact functions of several classifier genes (FAM26F, SAMD9L, GBP4, GIMAP5) have not yet been identified.

Table 4. Taqman RT-PCR confirmation of the Affymetrix microarray results

Taqman ID	Gene symbol	Gene name	Affymetrix ID	Sample groups	P < 0.05	ddCt
Hs00153304_m1	CD44	CD44 antigen	212014_x_at	AD vs normal	1.82E• 07	1.90
Hs00171022_m1	CXCL12	Chemokine (C-X-C motif) ligand 12	209687_at	AD vs normal	0.00305	• 2.04
				CRC vs normal	0.00735	• 1.95
Hs00179845_m1	MET	Met proto-oncogene	203510_at	AD vs normal	1.41E• 06	2.17
				CRC vs normal	0.00002	1.53
Hs00200350_m1	ABCA8	ATP-binding cassette, subfamily A (ABC1), member 8	204719_at	AD vs normal	0.000610	• 3.35
				CRC vs normal	0.00143	• 3.20
Hs00205545_m1	ADAMDEC1	ADAM-like, decysin 1	206134_at	AD vs normal	1.16E• 05	• 3.69
				CRC vs normal	9.18E• 05	• 2.74
Hs00214306_m1	TRPM6	Transient receptor potential cation channel, subfamily M, member 6	224412_s_at	AD vs normal	5.79E• 05	• 4.73
				UC vs normal	0.000385	• 4.63
Hs00153408_m1	MYC	v-myc myelocytomatosis viral oncogene homologue (avian)	202431_s_at	AD vs normal	5.99E• 06	2.35
Hs00171558_m1	TIMP1	Tissue inhibitor of metalloproteinase 1	201666_at	AD vs normal	3.90E• 07	2.58
				CRC vs normal	0.00153	2.74
				UC vs normal	0.000219	2.36
Hs00236937_m1	CXCL1	Chemokine (C-X-C motif) ligand 1	204470_at	CRC vs normal	0.0114	3.84
				UC vs normal	1.11E• 05	4.04
Hs00236966_m1	CXCL2	Chemokine (C-X-C motif) ligand 2	209774_x_at	CRC vs normal	0.00204	3.70
				UC vs normal	0.000592	3.68
Hs00266139_m1	CA1	Carbonic anhydrase I	205950_s_at	AD vs normal	0.000930	• 6.13
Hs00194353_m1	LCN2	Lipocalin 2	212531_at	AD vs normal	2.67E• 07	6.13
				CRC vs normal	0.000509	4.83
				UC vs normal	2.15E• 06	5.06
Hs00154230_m1	CALU	Calumenin	214845_s_at	CRC vs normal	0.0145	1.60
Hs00169795_m1	VWF	von Willebrand factor	202112_at	CR vs normal	0.55142	
				UC vs normal	0.000112	2.44
Hs00266237_m1	COL4A1	Collagen, type IV, a 1	211980_at	CRC vs normal	0.0283	3.36
Hs00156076_m1	BGN	Biglycan	213905_x_at	CRC vs normal	0.12042	
Hs00169777_m1	PECAM1	Platelet/endothelial cell adhesion molecule	208983_s_at	CRC vs normal	0.76378	
Hs00174103_m1	IL8	Interleukin 8	202859_x_at	CRC vs normal	0.0283	7.21
				UC vs normal	6.80E• 06	5.77
Hs00204187_m1	DUOX2	Dual oxidase 2	219727_at	UC vs normal	7.84E• 05	5.35
Hs00195812_m1	LIPG	Lipase, endothelial	219181_at	AD vs normal	0.000588	1.35
				CRC vs normal	0.00711	1.08
				UC vs normal	0.000588	1.35
Hs00829485_sH	IFITM2	IFN induced transmembrane protein 2 (1-8D)	201315_x_at	CRC vs normal	0.00114	2.26
Hs00171061_m1	CXCL3	Chemokine (C-X-C motif) ligand 3	207850_at	CRC vs normal	0.00384	3.22
				UC vs normal	7.48E• 05	3.58
Hs00277299_m1	IL1RN	Interleukin 1 receptor antagonist	212657_s_at	CRC vs normal	0.00714	4.66
				UC vs normal	1.10E• 05	5.30
Hs00234579_m1	MMP9	Matrix metalloproteinase 9	203936_s_at	UC vs normal	0.00724	1.85
Hs00160066_m1	PI3	Protease inhibitor 3, skin-derived (SKALP)	203691_at	UC vs normal	0.000257	4.26
Hs00197374_m1	UBD	Ubiquitin D	205890_s_at	UC vs normal	0.000261	3.20

Abbreviation: AD, adenoma.

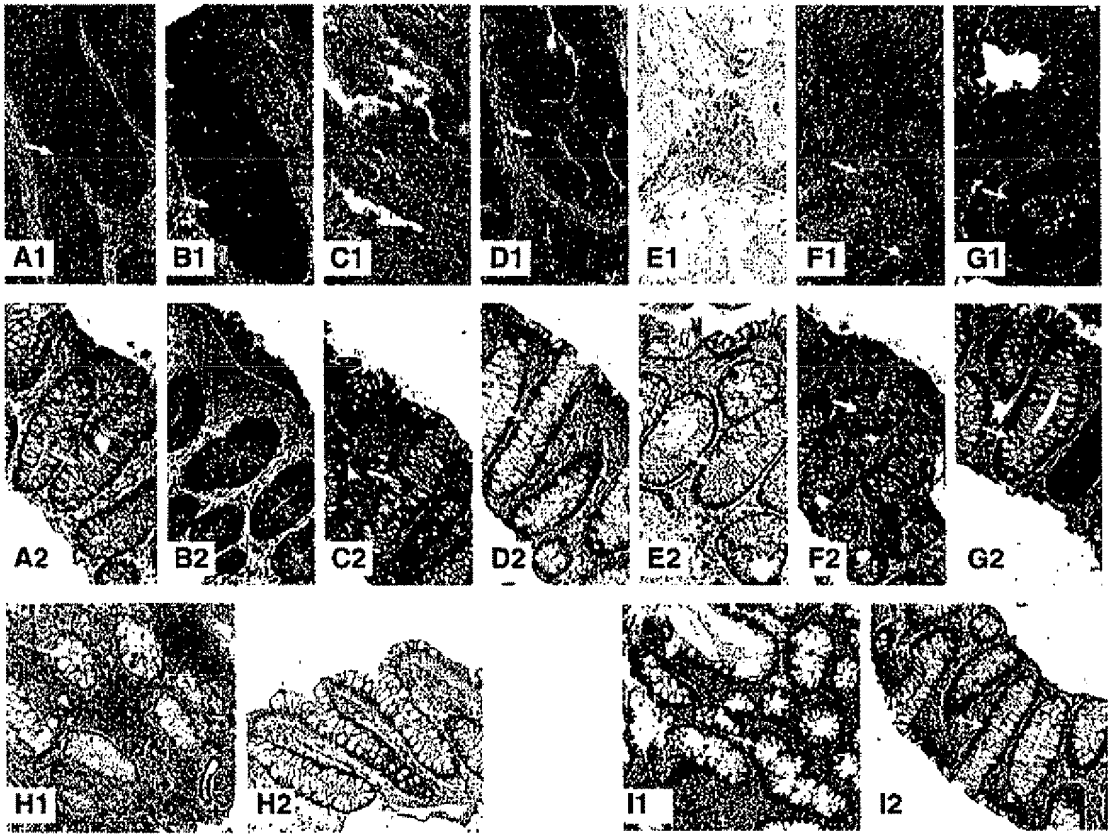


Figure 2. Immunostainings in TMA sections. A. Osteonectin immunostaining in CRC (A1) and healthy colonic mucosa (A2). B. Osteopontin immunostaining in CRC (B1) and healthy colonic mucosa (B2). C. Biglycan immunostaining in CRC (C1) and healthy colonic mucosa (C2). D. Collagen 4A1 immunostaining in CRC (D1) and healthy colonic mucosa (D2). E. von Willebrand factor immunostaining in CRC (E1) and healthy colonic mucosa (E2). F. MMP9 immunostaining in CRC (F1) and healthy colonic mucosa (F2). G. VEGF immunostaining in CRC (G1) and healthy colonic mucosa (G2). H. PECAM1 protein expression in active IBD (H1) and normal colonic tissue (H2). I. Collagen 4A1 protein expression in active IBD (I1) and healthy colonic mucosa (I2). The white arrows show the colonic epithelial cells. Elevated protein levels of osteonectin, osteopontin, biglycan, collagen 4a1, von Willenbrand factor, MMP9, and vascular endothelial growth factor were detected in CRC compared with healthy controls. In proportion to normal tissue, overexpression of PECAM1, and collagen 4a1 proteins was found in IBD.

Identification of Subclassifier Transcripts. The successful subdivision of IBD to ulcerative colitis and Crohn's disease was achieved by the expression of 58 genes such as cyclin G2, dual oxidase 2 and CEACAM7 (sensitivity 77.78%, specificity 100%). Adenomas with low-grade and high-grade dysplasia could be distinguished using 65 classifier probesets such as aquaporin 1, CXCL10, and complement factor 1 (sensitivity: 90.91%, specificity: 100%). Early and advanced stage colorectal carcinomas were differentiated by 34 discriminatory transcripts including transmembrane protein 37, interleukin 33, carbonic anhydrase 4, visinin-like 1, ubiquitous calcium-transporting ATPase, and CDK inhibitor 2B by high specificity (100%) and somewhat lower sensitivity values (66.67%; Table 2).

Expression of the Colorectal Cancer–Associated Tissue Markers in Peripheral Blood. The differentially expressed genes were determined by significance analysis of microarrays between colorectal cancer samples and healthy normal controls. The presence of these local tissue-specific mRNA expression markers in peripheral blood samples was also analyzed using the blood samples of 19 colorectal cancer and 11 healthy patients. Fifty-two transcripts were significantly up-regulated both in biopsy specimen and the peripheral blood of colorectal cancer patients compared with healthy normal controls. Three genes (SLC26A2 sulfate transporter, 227682_at, and UDP-glucose dehydrogenase) showed significantly decreased mRNA level both in colorectal cancer biopsy and blood samples compared with normals. In some colorectal cancer–related transcripts, mRNA expression in blood changed in the opposite way compared with their levels in cancer tissue. Seventeen genes showing elevated mRNA expression in colorectal

cancer biopsy samples were down-regulated in the peripheral blood of colorectal cancer patients, whereas 12 genes underexpressed in colorectal cancer tissue were found to be overexpressed in colorectal cancer blood samples (Table 3.).

Taqman RT-PCR Validation of 26 Selected Genes. The expression of all the 11 (6 up-regulated and 5 down-regulated in microarray analysis) adenoma-associated genes, 15 of the 18 colorectal cancer–related genes (15 overexpressed and 3 underexpressed), and all the 14 ulcerative colitis–associated genes (13 up-regulated and 1 down-regulated) correlated significantly with the Affymetrix results ($P < 0.05$). On average, the mRNA expression of 93% of the selected genes was verified by Taqman RT-PCR (Table 4.).

TMA Analysis and Blood Smear Immunocytochemistry Results. In accordance with mRNA expression results, elevated protein levels of osteonectin, osteopontin, biglycan, collagen 4a1, von Willenbrand factor, MMP9, and vascular endothelial growth factor were detected in colorectal cancer compared with healthy controls. Moderate cytoplasmatic osteopontin and osteonectin staining was found in the apical cytoplasm of epithelial cells in healthy colon tissue. Both osteonectin and osteopontin showed moderate to strong diffuse cytoplasmatic staining in colorectal cancer samples. Osteonectin protein expression was also significantly increased in blood smears of colorectal cancer patients (osteonectin positive mononuclear cells, 20.89% F 2.16%) compared with the normal (6.72% F 2.65%; $P = 6.35 \cdot 10^{-9}$; Supplementary Fig. S1). In colorectal cancer cases, strong subepithelial BGN immunostaining was found in lamina propria myofibroblast like cells and leukocytes. No epithelial BGN immunoreactivity was detected. Most of the normal samples were negative for BGN, but in some cases weak apical epithelial BGN immunostaining was found, and no subepithelial labeling was seen. Whereas all normal samples were negative for Col4A1, certain carcinomatous cells showed a moderate to strong epithelial Col4A1 immunostaining in colorectal cancer samples. There was no lamina propria immunoreactivity. Regarding vWF, there was moderate epithelial immunostaining in carcinomatous cells in colorectal cancer samples, and some vWF immunoreactivity was also seen scattered in the lamina propria whereas in normal cases no mucosal immunostaining was seen. Subepithelial MMP9 immunostaining was found to be moderate and strong in lamina propria leukocytes in colorectal cancer cases but not in carcinomatous epithelium. A diffuse weak intracytoplasmatic epithelial immunoreactivity was seen in normal samples. In case of vascular endothelial growth factor, epithelial immunoreactivity was found to be moderate to strong diffusely in carcinomatous cells of colorectal cancer samples. The subepithelium showed a moderate reaction. Weak to moderate subepithelial and luminal epithelial vascular endothelial growth factor immunoreactivity was found in almost all normal samples (Fig. 2).

In comparison with normal tissue, PECAM1 and collagen 4a1 proteins were overexpressed in IBD in accordance with the up-regulated mRNA levels detected by microarrays. In IBD samples there was a strong subepithelial PECAM1 immunoreaction in lamina propria leukocytoid cells. There was no epithelial immunoreaction in any of the normal samples. In several IBD samples a weak Col4a1 immunoreaction was found compared with normals. No subepithelial immunostaining could be detected (Fig. 2).

Discussion

In this study, 85 colonic biopsy samples and 30 peripheral blood samples were analyzed in total by whole genomic expression microarrays to identify local tissue classifiers between the diagnostic groups and to analyze the presence of the tissue expression markers in peripheral blood.

In the daily routine, the situation where the biopsy sample taken during the endoscopic intervention is not evaluable in the appropriate manner by conventional histology occurs relatively frequently. Diagnostic expression profile from the whole biopsy specimen can overcome the sampling error failures in histology.

For the objective, molecular-based classification of the biopsy samples into main diagnostic groups, classifier transcript sets were determined. Functional analysis of significant genes can provide important information, because with the identification of the main signaling pathways, the key genes characterizing the given pathomechanism can be found and used for diagnostic analysis.

Because an IBD, especially the long-standing ulcerative colitis, is a precancerous condition, the analysis of IBD specimen is important to find early the adenoma-dysplasia-carcinoma sequence–related genes.

All three IBD classifiers have been hypothesized to show increased expression in IBD. In case of a tissue injury associated with IBD, REG1A (regenerating islet-derived 1a) mRNA was observed to be highly expressed in colonic mucosa (38). The protein product of this gene has a positive regulatory effect on cell proliferation (39), and may contribute to reduce epithelial apoptosis in inflammation (38). Matrix metalloproteinase 3 (MMP3), involved in wound repair and tumor initiation, was also up-regulated in IBD (40). Microarray analysis done by Mizoguchi et al. indicated that the third classifier, chitinase 3-like 1 (CHI3L1) is overexpressed specifically in inflamed mucosa. CHI3L1 plays a pathogenic role in colitis, presumably by enhancing the adhesion and invasion of bacteria on/into colonic epithelial cells (41). Dysregulated host/microbial interactions seem to play a central role in the pathogenesis of IBD.

Analyzed by function, most of the colorectal cancer versus adenoma discriminatory genes are involved in intracellular signal transduction (GNG11, latrophilin, AKAP12, ELTD1, tensin 1, axin 2, GNB4, ELTD1), cell proliferation (IGFBP3, MCC, LATS2), cell adhesion (ROBO1, AEBP1, VWF, collagen 15A1, DDR2, PLEKHC1), and transcription regulation (like NR3C1, WWTR1, MEIS1, MEF2C, SNAI2). However, the functions of several discriminatory transcripts are still unknown. For instance, gremlin 1 (GREM1) which is represented among the classifiers with two probesets as an antagonist of BMP, may play a role in regulating organogenesis, body patterning, and tissue differentiation. It was overexpressed in various human tumors including carcinomas of the lung, ovary, kidney, breast, colon, pancreas, and sarcoma (42).

Polyps could be classified into adenomatous and hyperplastic polyps according to the expression levels of nine transcripts. The ABCA8 ABC transporter, which was previously found to be underexpressed in colorectal cancer (28, 43), showed decreased expression in adenoma compared with hyperplastic polyp samples. Lower glucagon mRNA levels in adenomas may refer to the altered intestinal barrier function (44) and disordered cell proliferation regulation. Interestingly, IGF1, overexpression of which is closely associated with the early stage of colorectal carcinogenesis (45), was found to be more intensely expressed in hyperplastic polyps than in adenomas. The lower peroxiredoxin 6 expression may indicate weaker protection against oxidative stress in adenomas. The exact functions of the MAMDC2, C2orf32, 229670_at, and KIAA1199 discriminatory transcripts have not yet been clarified.

The colorectal cancer versus IBD discriminatory genes are mainly immune and defense response–related genes (like CXCL9, CXCL10 chemokine ligands, CCR2, CCRL1 chemokine receptors, interleukin 18 binding protein, GBP1, GBP5, NOS2A, INDO, TNFSF13B, toll-like receptor 8, 227458_at) which showed decreased mRNA levels in colorectal cancer compared with IBD samples. CD38 expressed mainly in leukocytes is involved in cell adhesion and signal transduction, RARRES3 is a negative regulator of cell proliferation, whereas ECGF1 is a growth factor with angiogenic effects. RARRES3 has been reported to act as a tumor suppressor or growth regulator (46). Its decreased expression in colorectal cancer seems to support this assumption. Autocrine production of ECGF1 by endothelial cells may be a mechanism of inflammatory angiogenesis but not tumor angiogenesis and might be particularly important for the maintenance of damaged vasculature in IBD (47). The functions of some newly identified expression markers (FAM26F, FCRL5, SAMD9L, TNIP3) are unclarified.

The main diagnostic groups (colorectal cancer, IBD, adenomas, hyperplastic polyps) can be distinguished according to the mRNA expression levels of 18 genes determined by the random forest classification method with a 12.9% prediction error.

Besides the objective classification of the samples into main diagnostic groups, the differentiation among disease subtypes is also important for the improvement of the molecular-based diagnostics.

A relatively high number of classifiers is required for differentiation between high-grade and low-grade dysplastic villous adenomas. Several tumorigenesis-related discriminatory transcripts (such as HIPK1, CDC25B, CXCL2, and HMGA2) were found to be overexpressed in high-grade dysplastic adenoma referring to the high risk of colorectal cancer development (13, 14, 43, 48, 49). Homeodomain interacting protein kinase 1 (HIPK1) may thus play a role in tumorigenesis, perhaps by regulating the expression of p53 and/or Mdm2 (48). A correlation has previously been shown between the presence of HMGI proteins and the expression of a highly malignant phenotype in epithelial and fibroblastic rat thyroid cells. Moreover, HMGA2 seems to be involved in colorectal carcinogenesis (49).

Most of the colorectal cancer subtype classificators are involved in transport processes (calcium ion transport: transmembrane protein 37, ubiquitous calcium-transporting ATPase, CLIC6 chloride transporter, CYP4X1 electron transporter, GABRB2 chloride channel, SLC26A2 sulfate transporter), in metabolic processes (carbonic anhydrase 4, UDP glucuronosyltransferase 2 A3 polypeptide, glycosyltransferase-like 1B, monoacylglycerol O-acyltransferase 2), in cell adhesion and motility (espin, mucin-like protocadherin, tetraspanin 5), in signal transduction (visinin-like 1, C13orf18), and in cell cycle regulation (SPC25 kinetochore complex component, CDK inhibitor 2B). Visinin-like 1 (VSNL1) was overexpressed in neuroblastoma tumor specimens from patients with distant organ metastases compared with those without metastases (50). Decreased expression of the CDKN2B (alias p15) tumor suppressor gene is also typical in advanced colorectal cancer.

The future perspectives are to state the diagnosis and to perform screening using a more easily available sample source such as peripheral blood, and the further required diagnostic-therapeutic steps may be done with the help of them. However, WBC circulating in the peripheral blood tour all tissues of the body, and gene expression changes in them are affected by more conditions than the gene expression patterns in local tissue alterations. It is important to find the tissue markers that appear also in peripheral blood and can be specific for a given organic alteration.

Several colorectal cancer–associated tissue markers changed in peripheral blood parallel to the locally measured expression levels. Genes showing up-regulation in both biopsy and peripheral blood samples of colorectal cancer patients compared with normal controls are mainly involved in cell adhesion (like CD44, TGFb1, ICAM1, versican, collagen 18A1, pelota homologue endothelial cell adhesion molecule), cell proliferation (such as IFITM1, IFITM2, TIMP1, fascin homologue 1), and intracellular signal transduction (including S100A11, filamin A, and DDEF1), whereas the functions of nine transcripts (like CCDC85B, TM9SF4, C6orf145, and TMEM158) have not yet been identified. The gene signals may come from peripheral blood mononucleic cells, as well as from circulating tumor cells. Previously, we reported a significantly positive correlation between the number of circulating tumor cells and clinical properties of colorectal cancer (51). The underexpressed genes in both biopsy and blood samples are involved in metabolism (UGDH) and sulfate transport (SLC26A2), whereas the function of 227682_at is unknown. In some colorectal cancer–related transcripts, mRNA expression in blood changed in the opposite way compared with their levels in cancer tissue. This phenomenon may relate to secondary immunologic processes including tumor-infiltrating lymphocytes rather than circulating tumor cells.

The expression of selected IBD- and colorectal cancer–associated genes was also measured at protein level, on 292 tissue sections of 29 overlapping and 93 independent sets of patients. TMA technology allowed the standardized analysis of a large number of samples within a short time and the validation of some of the mRNA expression results. In accordance with mRNA expression results, elevated protein levels of osteonectin, osteopontin, biglycan, collagen 4A1, von Willenbrand factor, MMP9, and vascular endothelial growth factor were detected in colorectal cancer compared with healthy controls. Osteonectin protein expression in blood smears of colorectal cancer patients was also significantly

elevated compared with normal controls. Overexpression of PECAM1 and collagen 4a1 proteins was detected in IBD compared with normal tissue, in accordance with the up-regulated mRNA levels detected by microarray.

In conclusion, whole genomic microarray analysis using routine biopsy samples may be suitable for the identification of discriminative signatures for differential diagnostic purposes. Our results may serve as a basis of new gene expression pattern–based diagnostic methods like Taqman and/or LightCycler 480 real-time PCR cards. As the mRNA expression results showed a strong correlation with the protein level expression, simultaneous analysis of protein marker sets can also take place. Nowadays, antibodies recognizing a wide range of proteins in formol-paraffin tissue sections are available, offering immunostaining of disease-specific markers as a simple test for daily diagnostic utilization.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank László Nagy, M.D., Ph.D., and Beáta Scholtz, Ph.D., for their help with the Taqman real-time PCR analysis, and Gabriella Kónya for her work in preparing the TMA immunostainings.

References

1. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J Natl Cancer Inst 2004;96:1420–5.
2. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759–67.
3. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. Br J Surg 2002;89:845–60.
4. Hawkins NJ, Bariol C, Ward RL. The serrated neoplasia pathway. Pathology 2002;34:548–55.
5. Jass JR, Baker K, Zlobec I, et al. Advanced colorectal polyps with the molecular and morphological features of serrated polyps and adenomas: concept of a 'fusion' pathway to colorectal cancer. Histopathology 2006;49:121–31.
6. Kambara T, Simms LA, Whitehall VL, et al. BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. Gut 2004;53:1137–44.
7. Jass JR. Serrated adenoma of the colorectum and the DNA-methylator phenotype. Nat Clin Pract Oncol 2005;2:398–405.
8. Iino H, Jass JR, Simms LA, et al. DNA microsatellite instability in hyperplastic polyps, serrated adenomas, and mixed polyps: a mild mutator pathway for colorectal cancer? J Clin Pathol 1999;52:5–9.
9. Kwon HC, Kim SH, Roh MS, et al. Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. Dis Colon Rectum 2004;47:141–52.
10. Agrawal D, Chen T, Irby R, et al. Osteopontin identified as colon cancer tumor progression marker. C R Biol 2003;326:1041–3.
11. Bandres E, Cataian V, Sola I, et al. Dysregulation of apoptosis is a major mechanism in the lymph node involvement in colorectal carcinoma. J Oncol Rep 2004;12:287–92.
12. Bertucci F, Salas S, Eysteries S, et al. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene 2004;23:1377–91.
13. Birkenkamp-Demtroder K, Christansen LL, Olesen SH, et al. Gene expression in colorectal cancer. Cancer Res 2002;62:4352–63.
14. Williams NS, Gaynor RB, Scoggin S, et al. Identification and validation of genes involved in the pathogenesis of colorectal cancer using cDNA microarrays and RNA interference. Clin Cancer Res 2003;9:931–46.
15. Li M, Lin YM, Hasegawa S, et al. Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. Int J Oncol 2004;24:305–12.
16. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 2002;21:4120–8.
17. Yanagawa R, Furukawa Y, Tsunoda T, et al. Genome-wide screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray. Neoplasia 2001;3:395–401.
18. Frederiksen CM, Knudsen S, Laurberg S, Orntoft TF. Classification of Dukes' B and C colorectal cancers using expression arrays. J Cancer Res Clin Oncol 2003;129:263–71.
19. Zou TT, Selaru FM, Xu Y, Shustova V, Yin J, Mori Y. Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. Oncogene 2002;21:4855–62.
20. Birkenkamp-Demtroder K, Olesen SH, Sorensen FB, et al. Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid. Gut 2005;54:374–84.
21. Chiu ST, Hsieh FJ, Chen SW, Chen CL, Shu HF, Li H. Clinicopathologic correlation of up-regulated genes identified using cDNA microarray and real-time reverse transcription-PCR in human colorectal cancer. Cancer Epidemiol Biomarkers Prev 2005;14:437–43.
22. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 2004;22:1564–71.
23. Burczynski ME, Twine NC, Dukart G, et al. Transcriptional profiles in peripheral blood mononuclear cells prognostic of clinical outcomes in patients with advanced renal cell carcinoma. Clin Cancer Res 2005;11:1181–9.
24. Twine NC, Stover JA, Marshall B, et al. Disease-associated expression profiles in peripheral blood mononuclear cells from patients with advanced renal cell carcinoma. Cancer Res 2003;63:6069–75.
25. Martin KJ, Graner E, Li Y, et al. High-sensitivity array analysis of gene expression for the early detection of disseminated breast tumor cells in peripheral blood. Proc Natl Acad Sci U S A 2001;98:2646–51.
26. Smirnov DA, Zweitzig DR, Foulk BW, et al. Global gene expression profiling of circulating tumor cells. Cancer Res 2005;65:4993–7.
27. Toth K, Galamb O, Solymosi N, et al. Peripheral blood gene expression markers of colorectal diseases. Magy Beterv Arch 2007;60:531–9.
28. Galamb O, Gyorffy B, Sipos F, et al. Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature. Dis Markers 2008;25:1–16.
29. Van der Flier LG, Sabates-Bellver J, Oving I, et al. The intestinal Wnt/TCF signature. Gastroenterology 2007;132:628–32.
30. Csillag C, Nielsen OH, Vainer B, et al. Expression of the genes dual oxidase 2, lipocalin 2 and regenerating islet-derived 1 A in Crohn's disease. Scand J Gastroenterol 2007;42:454–63.
31. Gardina PJ, Clark TA, Shimada B, et al. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics 2006;7:325.
32. Borthakur A, Bhattacharyya S, Dudeja PK, Tobacman JK. Carrageenan induces interleukin-8 production through distinct Bcl10 pathway in normal human colonic epithelial cells. Am J Physiol Gastrointest Liver Physiol 2007;292:G829–38.
33. Mansilla F, Birkenkamp-Demtroder K, Kruhoffer M, et al. Differential expression of DHHC9 in microsatellite stable and instable human colorectal cancer subgroups. Br J Cancer 2007;96:1896–903.
34. Tumor Analysis Best Practices Working Group. Expression profiling-best practices for data generation and interpretation in clinical trials. Nat Rev Genet 2004;5:229–37.
35. Díaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics 2006;7:3.
36. Efron B, Tibshirani RJ. Improvements on cross-validation: the.632+ bootstrap method. J Am Stat Assoc 1997;92:548–60.
37. Friendly M. Visualizing Categorical Data. SAS Institute. Cary (NC): 2000.
38. Dieckgraefe BK, Crimmins DL, Landt V, et al. Expression of the regenerating gene family in inflammatory bowel disease mucosa: Reg IA upregulation, processing, and antiapoptotic activity. J Investig Med 2002;50:421–34.
39. Watanabe T, Yonekura H, Terazono K, Yamamoto H, Okamoto H. Complete nucleotide sequence of human reg gene and its expression in normal and tumoral tissues. The reg protein, pancreatic stone protein, and pancreatic thread protein are one and the same product of the gene. J Biol Chem 1990;265:7432–9.

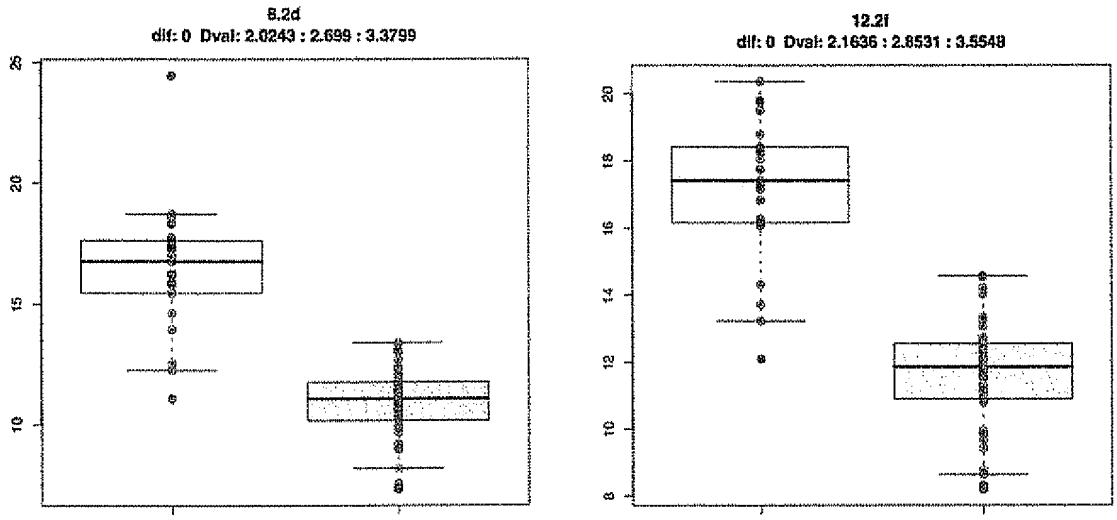
40. Lawrance IC, Fiocchi C, Chakravarti S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. Hum Mol Genet 2001;10:445–56.
41. Mizoguchi E. Chitinase 3-like-1 exacerbates intestinal inflammation by enhancing bacterial adhesion and invasion in colonic epithelial cells. Gastroenterology 2006;130:398–411.
42. Namkoong H, Shin SM, Kim HK, et al. The bone morphogenetic protein antagonist gremlin 1 is overexpressed in human cancers and interacts with YWHAH protein. BMC Cancer 2006;6:74.
43. Croner RS, Foertsch T, Brueckl WM, et al. Common denominator genes that distinguish colorectal carcinoma from normal mucosa. Int J Colorectal Dis 2005;20:353–62.
44. Benjamin MA, McKay DM, Yang PC, Cameron H, Perdue MH. Glucagon-like peptide-2 enhances intestinal epithelial barrier function of both transcellular and paracellular pathways in the mouse. Gut 2000;47:112–9.
45. Nosho K, Yamamoto H, Taniguchi H, et al. Interplay of insulin-like growth factor-II, insulin-like growth factor-I, insulin-like growth factor-I receptor, COX-2, and matrix metalloproteinase-7, play key roles in the early stage of colorectal carcinogenesis. Clin Cancer Res 2004;10:7950–7.
46. Jiang SY, Chou JM, Lou FJ, et al. Decreased expression of type II tumor suppressor gene RARRES3 in tissues of hepatocellular carcinoma and cholangiocarcinoma. World J Gastroenterol 2005;11: 948–53.
47. Saito S, Tsuno NH, Sunami E, et al. Expression of platelet-derived endothelial cell growth factor in inflammatory bowel disease. J Gastroentarol 2003;38:229–37.
48. Kondo S, Lu Y, Debbas M, et al. Characterization of cells and gene-targeted mice deficient for the p53-binding kinase homeodomain-interacting protein kinase 1 (HIPK1). Proc Natl Acad Sci U S A 2003; 100:5431–6.
49. Fedele M, Bandiera A, Chiappetta G, et al. Human colorectal carcinomas express high levels of high mobility group HMGI(Y) proteins. Cancer Res 1996;56:1896–901.
50. Xie Y, Chan H, Fan J, et al. Involvement of visinin-like protein-1 (VSNL-1) in regulating proliferative and invasive properties of neuroblastoma. Carcinogenesis 2007;28:2122–30.
51. Molnar B, Floro L, Sipos F, Toth B, Sreter L, Tulassay Z. Elevation in peripheral blood circulating tumor cell number correlates with macroscopic progression in UICC stage IV colorectal cancer patients. Dis Markers 2008;24:141–50.

EXHIBIT 5

FIGURE 4

A

B



DETECTION OF SEQID_7 RNA BY RT-PCR

A quantitative TaqMan assay was developed to evaluate the relative expression of the SEQID_7 sequence derived from the discovery of clones 8.2d (A) and 12.2f (B). The expression of SEQID 7 was measured in in 20 normal (blue) and 51 adenomas (red) colorectal tissue specimens using the oligonucleotide primers, 5'- CAGACTTTACATCATGGGTGACCA (forward) and 5'- GCCATCCTGTGGCCCC (reverse), together with a TaqMan probe, TCCCGCAGAGTTGTACAGAACCTCCC, targeting a 162nt segment located from nucleotides 3364 to 3432 . Resulting expression data were normalised relatively to HPRT1 using the Livak method (2-ΔΔCT) and are displayed as box-and-whisker diagrams, where the 'boxes' give the median values and the lower and upper quartiles of the datasets.

FIGURE 5

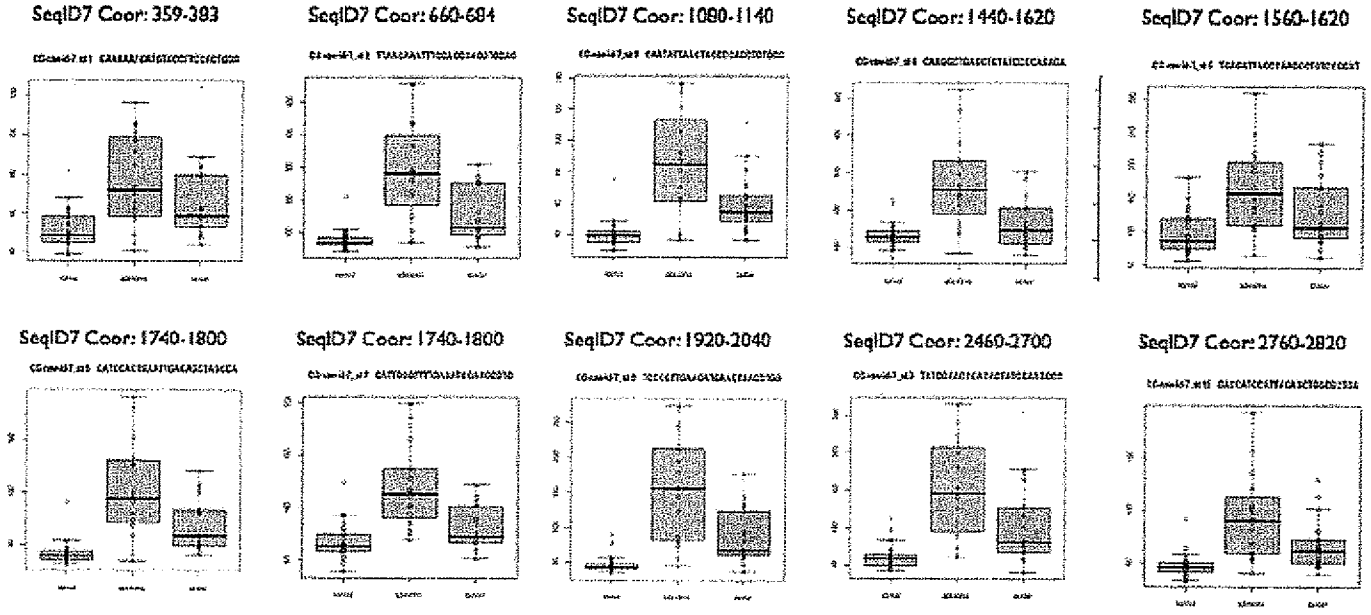
DEMONSTRATION OF DIAGNOSTIC UTILITY OF SEQID_7.

(A) Eleven oligonucleotide probes were designed to monitor the expression of SEQ ID 7 RNA in colon tissue specimens from 30 normal subjects (left bars) and 21 subjects with either adenomas (middle bars) or colorectal cancer (right bars) using a custom-made microarray gene chip. The CG Custom GeneChips were processed according to manufacturer's instructions for the Affymetrix HuGene ST 1.0 array using biotinylated DNA derived from an original primary RNA concentration of 100ng. Hybridization to the Custom Chip CG_AGPa520460F was carried out at 45°C for 16-18 hours. Finally, the chips were washed, stained and scanned as recommended by Affymetrix using an Affymetrix Scanner 3000. Transcript expression levels were calculated by both Microarray Suite (MAS) 5.0 (Affymetrix) and the Robust Multichip Average (RMA) normalization techniques (Affymetrix. GeneChip expression data analysis fundamentals. Affymetrix, Santa Clara, CA USA, 2001; Hubbell et al. Bioinformatics, 18:1585–1592, 2002; Irizarry et al. Nucleic Acid Research, 31, 2003) MAS normalized data was used for performing standard quality control routines and the final data set was normalized with RMA for all subsequent analyses.

(B) The average expression across the eleven oligonucleotide probes described in (A).

FIGURE 5

A


SeqID7 Coor: 359-393
SeqID7 Coor: 660-684
SeqID7 Coor: 1080-1140
SeqID7 Coor: 1440-1620
SeqID7 Coor: 1560-1620
SeqID7 Coor: 1740-1800
SeqID7 Coor: 1740-1800
SeqID7 Coor: 1920-2040
SeqID7 Coor: 2460-2700
SeqID7 Coor: 2760-2820

B

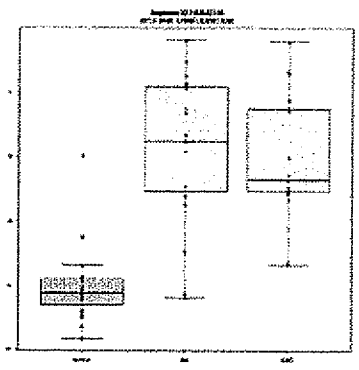


EXHIBIT B

Transcriptome Profile of Human Colorectal Adenomas

Jacob Sabates-Bellver,[1] Laurens G. Van der Flier,[4] Mariagrazia de Palo,[5] Elisa Cattaneo,[1] Caroline Maake,[2] Hubert Rehrauer,[3] Endre Laczko,[3] Michal A. Kurowski,[9] Janusz M. Bujnicki,[9] Mirco Menigatti,[7] Judith Luz,[8] Teresa V. Ranalli,[6] Vito Gomes,[6] Alfredo Pastorelli,[5] Roberto Faggiani,[5] Marcello Anti,[5] Josef Jiricny,[1] Hans Clevers,[4] and Giancarlo Marra[1]

[1]Institute of Molecular Cancer Research, [2]Institute of Anatomy, and [3]Functional Genomic Center, University of Zurich, Zurich, Switzerland; [4]Hubrecht Institute, Netherlands Institute for Developmental Biology, Utrecht, the Netherlands; [5]Gastroenterology Unit and [6]Pathology Department, Belcolle Hospital, Viterbo, Italy; [7]Institute of Biochemistry and Genetics and [8]Division of Medical Genetics, University of Basel, Basel, Switzerland; and [9]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warsaw, Poland

Abstract

Colorectal cancers are believed to arise predominantly from adenomas. Although these precancerous lesions have been subjected to extensive clinical, pathologic, and molecular analyses, little is currently known about the global gene expression changes accompanying their formation. To characterize the molecular processes underlying the transformation of normal colonic epithelium, we compared the transcriptomes of 32 prospectively collected adenomas with those of normal mucosa from the same individuals. Important differences emerged not only between the expression profiles of normal and adenomatous tissues but also between those of small and large adenomas. A key feature of the transformation process was the remodeling of the Wnt pathway reflected in patent overexpression and underexpression of 78 known components of this signaling cascade. The expression of 19 Wnt targets was closely correlated with clear up-regulation of KIAA1199, whose function is currently unknown. In normal mucosa, KIAA1199 expression was confined to cells in the lower portion of intestinal crypts, where Wnt signaling is physiologically active, but it was markedly increased in all adenomas, where it was expressed in most of the epithelial cells, and in colon cancer cell lines, it was markedly reduced by inactivation of the β-catenin/T-cell factor(s) transcription complex, the pivotal mediator of Wnt signaling. Our transcriptomic profiles of normal colonic mucosa and colorectal adenomas shed new light on the early stages of colorectal tumorigenesis and identified KIAA1199 as a novel target of the Wnt signaling pathway and a putative marker of colorectal adenomatous transformation.
(Mol Cancer Res 2007;5(12):1263–75)

Introduction

In developed countries, sporadic adenomatous colorectal polyps are found in roughly one third of asymptomatic adults below the age of 50 who undergo colonoscopy. Depending on their characteristics (multiplicity, size, histologic features, and degree of dysplasia), these lesions can be associated with a substantial risk of recurrence (up to 60% at 3 years) and the development of advanced neoplastic disease (reviewed in ref. 1 and references therein). It has been estimated that 15% of all adenomas measuring ≥1 cm will progress to carcinomas within 10 years of their detection (2).

Although adenomatous polyps are not the only precancerous lesions in the colorectum, they are the most common, and they are the precursors of most of the cancers in this organ. In these neoplasms, the transformation process begins in the epithelial crypts and seems to result from qualitative, quantitative, and spatial subversion of the Wnt signaling pathway, the physiologic regulator of epithelial homeostasis (3-5). This adenoma-carcinoma pathway of tumorigenesis is characterized by mutations involving various components of this pathway (e.g., APC, whose germ-line mutations are responsible for familial adenomatous polyposis; CTNNB1, which encodes a subunit of the cadherin protein complex known as β-catenin; and Axin, the gene encoding a multidomain scaffold protein that is essential for β-catenin degradation). The result of these mutations is an accumulation of β-catenin, first in the cytoplasm and then in the nucleus, where it associates with DNA-binding proteins of the T-cell factor (TCF)/lymphoid enhancer factor family, transforming them from transcriptional repressors into transcriptional activators that affect the expression of numerous genes involved in epithelial homeostasis.

Although the key role played by adenomatous polyps in colorectal tumorigenesis is widely acknowledged, the gene expression changes that trigger or accompany their development

Received 6/9/07; revised 7/28/07; accepted 8/2/07.
Grant support: Zurich Cancer League and Swiss National Science Foundation.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Note: Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).
J. Sabates-Bellver and L.G. Van der Flier contributed equally to this work.
Requests for reprints: Giancarlo Marra, Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. Phone: 41-044-635-3472; Fax: 41-044-635-3484. E-mail: marra@imcr.uzh.ch

doi:10.1158/1541-7786.MCR-07-0267

have never been comprehensively studied. We therefore conducted a transcriptomic analysis of prospectively collected colorectal adenomas using a standardized oligonucleotide microarray covering the entire human genome. This study not only provided new information that is fundamental for future molecular characterization of these precancerous lesions but also allowed us to identify a putative marker of colorectal tumorigenesis.

Results

The focus of our study was the adenoma-adenocarcinoma pathway of colorectal carcinogenesis, which is closely linked to deregulation of the Wnt signaling pathway. To gain insight into the early steps of this process, we confined our investigation exclusively to sporadic, pedunculated colorectal adenomas (type 0-Ip of the Paris classification; ref. 6). Nonpolypoid and sessile polypoid lesions were not included because in some cases their transformation is believed to proceed along nonadenomatous pathways (7). Details on our case selection criteria are provided in Materials and Methods.

Thirty-two pedunculated adenomatous polyps, each with matched samples of normal mucosa, were prospectively collected from 28 patients (Table 1). The total number of synchronous and previously excised adenomas was <3 in 18 of 28 patients and 3 to 15 in the remaining 10. In this latter subgroup, the absence of APC- or MYH-associated multiple adenomatosis had been confirmed by genetic testing of lymphocyte DNA. Histologic analysis of one polyp (case NM) revealed superficial infiltration of the submucosa, but this case was not excluded because the region sampled for microarray analysis was clearly adenomatous. (As noted below, this finding was consistent with the results of hierarchical cluster analysis shown in Supplementary Fig. S1.).

Analysis of microarray data for the 32 adenoma/normal mucosa tissue pairs revealed that 31,033 of the probes were expressed in one or both of the tissue groups. The normal tissues were effectively segregated from the adenomas in four unsupervised analyses of the expression levels of these genes [hierarchical clustering, principal component analysis (PCA), correlation analysis, and correspondence analysis (CA); see Materials and Methods for details; Fig. 1]. In a separate

Table 1. Characteristics of the 28 Patients with Adenomatous Polyps Included in the Study

Patient	Age (y)	Sex	Colon segment involved	Maximum adenoma diameter (mm)	Microscopic appearance	Highest degree of dysplasia in the adenoma*	Degree of dysplasia at sampling site*	No. adenomas at study colonoscopy[c]	No. previously excised adenomas[b]	No. previous and/or synchronous hyperplastic polyps	Familiarity for colorectal cancer (relative, onset age)
GL[x]	49	M	D/S	10/10	T-V/T-V	H/L	H/L	9	4[k,f]	15**	Mother, 70
PR	74	F	S	20	T-V	H	H	2	—[cc]	0	No
PC[x]	69	M	S/S	10/20	T/T-V	H/H	H/H	10	—	1	No
FP[x]	57	M	S/S	15/30	T/T-V	H/H	H/L	1	—	1	Mother, 69
CD[x]	71	M	T/R	15/10	T/T	H/L	H/L	2	7[k]	2	No
MA	65	M	R	15	T-V	L	L	2	—	0	No
ME	63	M	R	15	T-V	L	L	9	—	1	Father, 79
RA	64	F	A	15	T	L	L	1	7[k]	0	Sister, 68
PR	72	M	R	40	T-V	H	H	5	—	0	No
SD	56	M	A	15	T	H	H	1	1[k]	0	Mother, 83; sister, 87
MP	38	M	S	15	T-V	L	L	2	—	0	Father, 79
MP	61	M	S	20	T	L	L	3	—	2	no
LG	41	M	R	20	T-V	H	L	5 (2 serrated)	—	0	Father, 60
LS	45	M	S	20	T	H	H	1	—	1	Father, 60
BG	58	M	D	15	T-V	L	L	2	—	1	No
PL	69	F	S	15	T-V	L	L	2	—	0	No
SMA	52	F	S	30	T	H	H	2	—	1	No
MR	58	F	D	20	T	L	L	2	—	0	No
GN	69	M	R	40	T	H	H	2	—	0	No
BA	59	M	S	30	T	L	L	6	—	0	No
PF	56	M	S	30	T	L	L	2	—	0	No
RC	55	F	A	30	T-V	L	L	12	3[k,f]	1	No
TMA	58	F	S	10	T	L	L	1	—	0	Mother, 85
NM	52	M	R	35	T-V	T1[bb]	H	1	—	0	No
MA	83	M	S	10	T-V	H	H	2	—	1	No
MM	50	M	S	30	T-V	H	H	2	—	0	Father's brother, 65
NF	79	M	A	20	T-V	L	L	2	—	0	Mother, 70
PN	67	F	S	15	T-V	L	L	1	—	1	No

Abbreviations: M, male; F, female; A, ascending colon; T, transversum; D, descending colon; S, sigmoid colon; R, rectum; T, tubular; T-V, tubulovillous; L, low-grade dysplasia; H, high-grade dysplasia.

*Low-grade versus high-grade dysplasia as defined by the WHO classification of tumors of the digestive system, editorial and consensus conference in Lyon, France, November 6-9, 1999, IARC.

cThis number includes the adenoma(s) subjected to microarray analysis.

bTotal number of adenomas detected and excised during previous colonoscopies.

xTwo adenomas from these patients were analyzed.

kThese cases were considered as recurrent adenomas for the CCA.

(The index colonoscopy was done in a different center about 10 y before the study colonoscopy.

**Hyperplastic polyposis.

ccNo previous colonoscopies.

bbSuperficial submucosal invasion (T1). The tissue collected for microarray came from the adenomatous portion of the polyp.

analysis, these two tissue groups were also unequivocally distinguished from a previously described set of 25 colon cancers (8), which we reanalyzed for this study with the same microarray used to characterize the adenomas and normal mucosa (Supplementary Fig. S1).

Almost half of the expressed probes (15,059 of 31,033) displayed significant expression changes in adenomas. Those with fold changes ≥2 (1,190 probes up-regulated and 2,469 down-regulated in adenomas) were subjected to gene ontology analysis to identify the biological processes involved in the transition of normal mucosa to adenoma. The most significant results of this analysis are listed in Supplementary Table S1. The processes that were most markedly overrepresented among genes that were up-regulated in adenomas included mitosis, DNA replication, and spindle organization. Down-regulated genes were predominantly involved in host immune defense, inorganic anion transport, organ development, and inflammatory response, although a small group of genes involved in the latter process was up-regulated in adenomas (Supplementary Fig. S2).

We then analyzed the transcript levels of 319 genes believed to be components of the complex Wnt signaling pathway (Supplementary Table S2). Sixty-six of these genes (21%) were not expressed in either the normal or adenomatous tissue, and 34% were expressed similarly in both tissue groups. The remaining 144 genes displayed significantly altered expression in adenomas, and 78 of 144 displayed fold changes of ≥2.

A supervised extension of CA (9), canonical CA (CCA), was then used to identify possible correlations between gene expression patterns and clinical or pathologic variables. Four of the variables considered (adenoma diameter, colon segment of origin, degree of dysplasia, and adenoma recurrence; see Table 1) were clearly associated with distinct clusters of expression profiles (Fig. 2, variables in A and clusters for adenoma diameter in B; more details in the legend to this figure). The profile of adenomas measuring >20 mm could be easily distinguished from those of smaller (≤20 mm) adenomas. As shown by CCA and visualized on the corresponding CCA score plot (Fig. 2B), the centers of the three adenoma size clusters are distributed along the principal CCA axis (the vertical axis in Fig. 2B, the most important axis of separation of the expression profiles) in a definite order, with increasing diameters corresponding to progressively higher CCA scores. The variable large adenoma diameter was closely correlated with the vertical CCA axis (i.e., its vector "d>20mm" in Fig. 2A is almost parallel to this axis). It is interesting to note that the same correlation can be observed for the variable high-degree dysplasia (i.e., represented in Fig. 2A by vector "Hd"). This finding confirms the expected correlation between larger diameters and higher-degree dysplasia.

The CCA plot of the 11,709 modeled probes (loading plot, not shown) suggested that the distinction between the three size groups of adenomas is due to a complex network of relatively small changes in the expression of numerous genes (as opposed to marked changes involving a limited number of genes). Nevertheless, to maximize the use of the extensive data sets, we selected the 500 probes with the highest loading scores along the CCA axis 1 and isolated a set of genes whose expression changes displayed significant positive or negative correlation with adenoma size (Supplementary Table S3). Although their association with adenomas must be validated in a larger series, these are the expression changes most likely to play causal roles in the progression of these tumors.

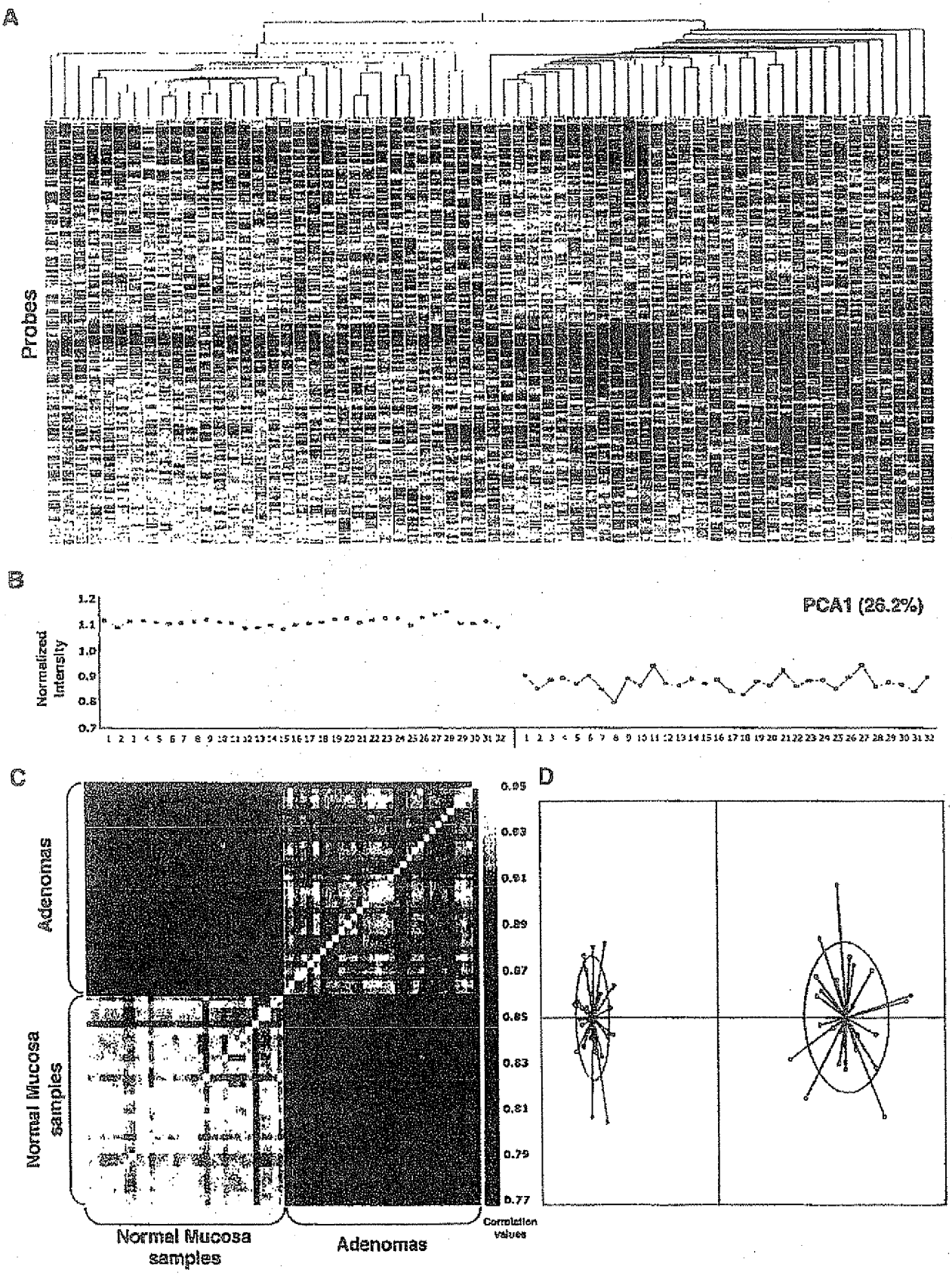
It should be mentioned that normal mucosa from the sigmoid colon had an expression profile that differed significantly from that of tissues from other colon segments (Fig. 2A). This finding will be explored in a future study conducted on a large series of normal mucosa samples from different colorectal segments.

The transcriptional profile of the 32 adenomas was thoroughly analyzed to identify genes likely to be involved in the development and evolution of these lesions. One of the first features that attracted our attention was the marked up-regulation of KIAA1199 (Supplementary Table S4), a gene encoding a protein with unknown function. Its overexpression was striking in all colorectal adenomas we examined (average increases of 54.8-fold compared with normal mucosa) and in carcinomas (8). These findings were fully confirmed by real-time reverse transcription-PCR analysis of RNA extracted from samples used for the microarray study and from additional samples collected after the present study was completed (Supplementary Fig. S3).

In light of these findings, it was natural to wonder whether KIAA1199 might be a novel positively regulated target of Wnt signaling, which is characteristically deregulated in colorectal tumors. Previous microarray studies indicated that genes coregulated at the transcriptional level under different conditions tend to be involved in the same processes and pathways, and the analysis of transcriptional coexpression has been used to predict the function of novel genes (10-12). Therefore, we conducted a search for known Wnt targets (listed in Supplementary Table S5) among the genes whose expression patterns in all the tissue samples significantly correlated with those of KIAA1199. (The procedure used in this analysis is summarized in Materials and Methods and Supplementary Fig. S4.) Forty-nine percent of the known Wnt targets that were overexpressed in our adenoma samples had expression patterns that were positively correlated with that of KIAA1199 (Fig. 3A and B) as opposed to only 7.9% of the overexpressed genes that are not considered Wnt targets ($P < 0.0001$).

Evidence of the potential involvement of KIAA1199 in the Wnt signaling pathway had also emerged from another study by our group (13). A combined analysis of microarray data of tissues and cell lines placed KIAA1199 at the top of a list of genes [Supplementary Table S1 of ref. 13] that were up-regulated in colorectal adenomas and down-regulated in colon cancer cell lines that had undergone stable transfection with doxycycline-inducible forms of dominant-negative TCF1 or TCF4 to suppress Wnt signaling (14, 15). In the present study, KIAA1199 was also found to be markedly down-regulated in LS174T colon cancer cells in which Wnt signaling had been blocked by the induction of h-catenin small interfering RNA or NH_2-terminal–deleted TCF4 (15, 16). The dramatic decrease in KIAA1199 mRNA levels associated with this inhibition of the Wnt pathway was confirmed by Northern blotting (Fig. 3C).

In general, Wnt target genes are expressed predominantly in the proliferating compartment of normal intestinal crypts (lower portion), and their expression is appreciably increased in


A
Probes
B
Normalized Intensity
PCA1 (26.2%)
C
Adenomas
Normal Mucosa samples
Normal Mucosa samples
Adenomas
Correlation values
D

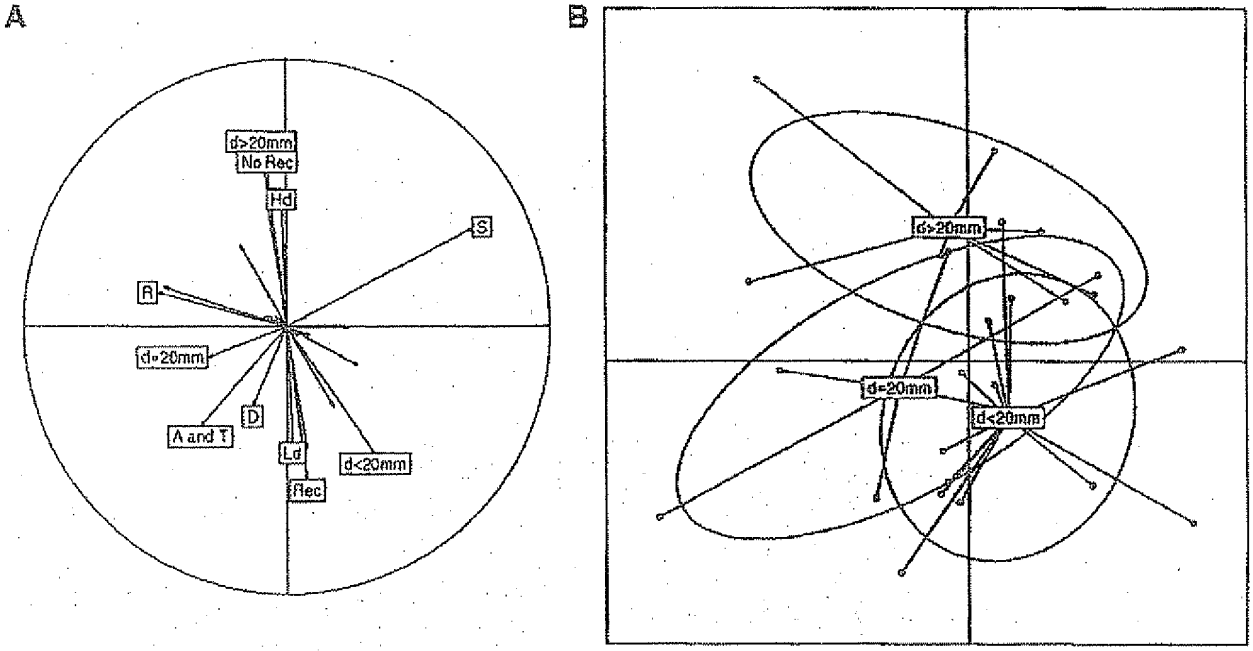


FIGURE 2. Clinical/pathologic variables that correlate with distinct gene expression profiles. The panels summarize the most important results of the CCA of mRNA intensity log-ratio values (adenoma: normal) of expressed genes. For clarity, CCA axis 1 has been drawn vertically in both panels. A. Correlation between specific clinical/pathologic variables (adenoma diameter, colon segment of origin, degree of dysplasia, and adenoma recurrence) and clusters of differential gene expression profiles (coded as log-ratio profiles), such as those shown in B. Each vector represents a specific value for a given variable (e.g., adenoma diameter of >20 mm and high-degree dysplasia) and points toward the center of the profile cluster correlated with the clinical/pathologic characteristic it represents. If the centers for each specific value are separated, the corresponding vectors point in distinct directions; otherwise, they are directed toward the same point. In the former case, the represented variable can be assumed to be significantly correlated with the profiles; in the latter case, there is no correlation. The length of the vector reflects the strength of the correlation: those approaching the circumference of the correlation circle, which represents a correlation value of 1, indicate stronger correlation than shorter vectors (correlation closer to 0). d, diameter; Hd, high-degree dysplasia; Ld, low-degree dysplasia; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum; Rec, recurrent adenomas; no Rec, no recurrent adenomas. Unlabeled vectors are related to variables that were not clearly associated with any distinct cluster of expression profiles. Larger adenomas were predictably associated with high-degree dysplasia. In contrast, their association with nonrecurrence was unexpected and probably due to the fact that patients who had already undergone endoscopic polypectomy (i.e., those with recurrence) presented relatively recent-onset (consequently, smaller) polyps at the study colonoscopy. B. CCA score plot with samples grouped by adenoma diameter. Each of the three size-related groups is delimited by an ellipse with the center labeled. The ellipse representing the adenomas measuring >20 mm in diameter shows very little overlap with those of the other two groups (adenomas with diameters of 20 mm and those with diameters of <20 mm).

adenomatous glands (15). Our analysis of human tissues with preserved architecture indicated that these are also attributes of KIAA1199. In in situ hybridization studies, KIAA1199 mRNA was detectable only in the lower portion of normal colonic epithelial crypts (Fig. 4A and B), and its expression levels were much higher in dysplastic glands (Fig. 4C). These patterns were confirmed at the protein level by immunohistochemistry done with an antibody raised in our laboratory (Fig. 4D-J). This analysis also revealed that the KIAA1199 is a cytoplasmic protein whose expression is most intense near the cell membrane, particularly on the luminal side of the dysplastic cell multilayer (Fig. 4F-J).

FIGURE 1. Unsupervised analyses of microarray data. A. Hierarchical clustering analysis. The 64 tissue samples represented on the X axis include 32 normal mucosal samples (green branches) and 32 adenomas (red branches). Each probe plotted on the Y axis is color coded to indicate the level of expression of the gene relative to its median expression level across the entire tissue sample set (blue, low; red, high). In the adenoma dendrogram, branches representing individual samples and small groups merge at higher levels than those of the normal mucosa dendrogram, reflecting lower-level correlation (i.e., higher variability among the adenoma specimens). B. PCA. Profile plot of the normalized first principal component (PCA1) across the 64 specimens (green dots, normal mucosa; red dots, adenomas). The two tissue groups differ significantly in terms of PCA1 ($P < 0.0001$), which accounted for 26% of the total variance. Note the higher variability of the PCA1 values in the adenoma group (higher fluctuation). C. Correlation analysis. Tile plot visualization of the pairwise correlations of the samples. Correlation values are indicated on the grayscale column (white > black: high > low). High correlation is observed among the samples within each group (top right quadrant, adenomas; bottom left quadrant, normal mucosa), although the adenomas displayed somewhat greater diversity (i.e., on the whole, the gray tones in the top right quadrant are darker than those in the bottom left quadrant). Top left and bottom right quadrants, normal and adenoma samples are poorly correlated. However, samples from the same patient generally showed higher correlation than that observed between normal and adenoma samples from different patients (bright pixels on the secondary diagonals in the top left and bottom right quadrants). This finding probably reflects the strong influence of several factors, including the individual genetic background and lifestyle and the fact that the normal and adenomatous tissues from a given patient were from the same colon segment. D. CA of mRNA log(intensity) values of expressed genes from 27 of the 32 tissue pairs (green dots, normal mucosa; red dots, adenoma). The other five pairs were excluded from this analysis because one of the two samples behaved as an outlier. Limiting our analysis to the more homogeneous pairs facilitated the comparison of the gene expression profiles for the two tissue groups and allowed more reliable identification of clinical/pathologic variables associated with profile scatter (see Fig. 2). The areas delimited by the ellipses represent 95% of the estimated binormal distribution of the sample scores on the first and second CA axes. The map of the sample scores on the first two axes shows that CA efficiently discriminates between normal and adenoma samples. Higher variability is evident in the adenoma group, where the samples are more widely dispersed.

Discussion

Adenomatous colorectal polyps are one of the most common human tumors and the most frequent precancerous lesions in the colorectum, but their transcriptome has been only partially analyzed, and the data are generally based on a limited number of cases (17-20). We attempted to fill this gap by doing a comprehensive whole-genome microarray analysis of a large, highly homogenous set of adenomas that was collected prospectively.

A comparison of the transcriptomes of adenomatous polyps and segment-matched samples of normal colorectal mucosa revealed evidence of broad-scale remodeling. As a starting point for future verification studies, we have drawn up a list of 478 genes that were significantly up-regulated (n = 153) or down-regulated (n = 325) in the adenomatous tissues (fold changes of ≥4; Supplementary Table S4). Space constraints preclude more than a cursory examination of this list, but we have highlighted in Table 2 certain aspects that we feel are particularly interesting in terms of their relevance to the process of adenoma formation. For instance, transcription regulation seems to be extensively modified. Twenty-nine molecules involved in this process were expressed in adenomas at levels

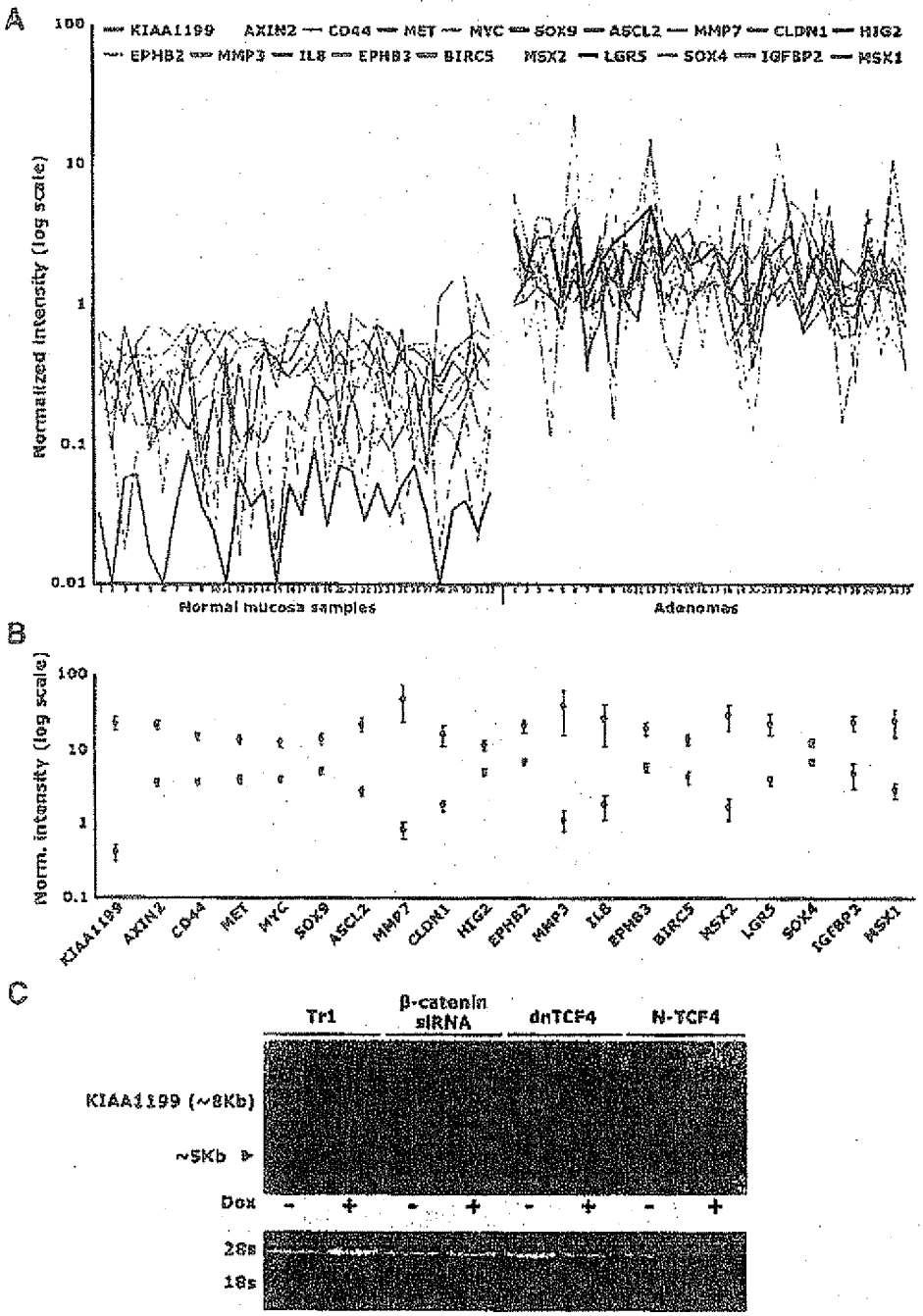


FIGURE 3. KIAA1199 is a putative target of Wnt signaling. A, Degree of correlation between the expression of KIAA1199 mRNA and that of 19 known Wnt signaling target genes identified with the procedure described in Materials and Methods, Results, and Supplementary Fig. S4. For each of the 20 genes, the graph shows the normalized intensity of expression level (plotted on the Y axis) in each of the 32 adenomas and corresponding samples of normal mucosa (X axis). B, Mean expression of each gene in normal mucosa (green dots) and adenomas (red dots). Bars, confidence interval. C, Northern blot showing reduced KIAA1199 expression in LS174T cells following doxycycline-mediated induction of h-catenin small interfering RNA, dominant-negative TCF4 (dnTCF4), or NH_2-terminal–deleted TCF4 (N-TCF4). The ~8-kb band corresponds to full-length KIAA1199 mRNA. The lower band (~5 kb) may represent an alternative form of this mRNA. Dox, cell transfectants grown in the presence or absence of doxycycline; Tr1, a parental clone (i.e., cells expressing the repressor protein modified by doxycycline but not transfected with h-catenin small interfering RNA, dominant-negative TCF4, or NH_2-terminal–deleted TCF4) used as a control of doxycycline exposure. Bottom, ethidium bromide–stained agarose gel as a loading control.

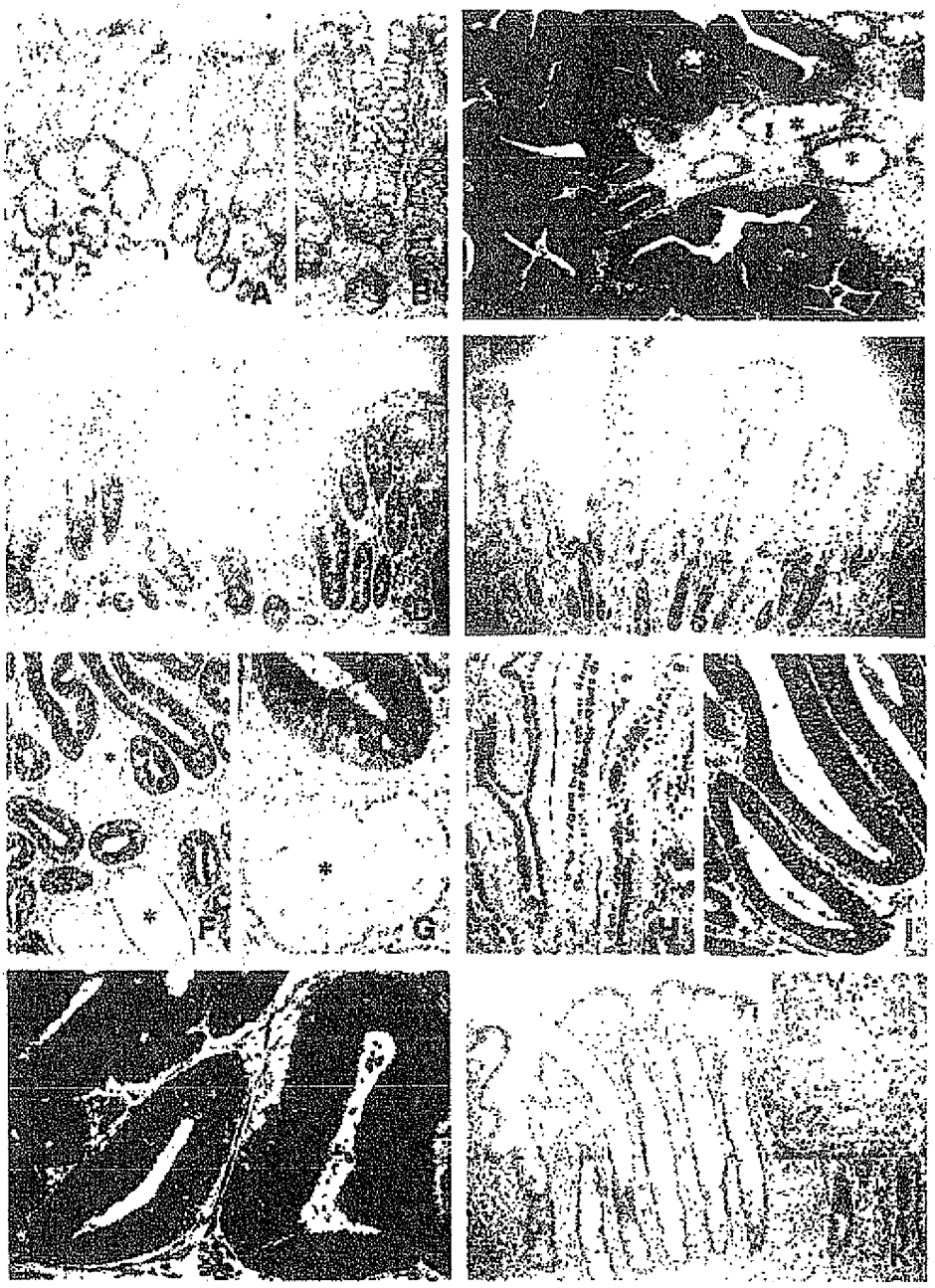


FIGURE 4. Expression of KIAA1199 mRNA and protein in normal intestinal mucosa and colorectal tumors. In situ hybridization studies (A-C) localized KIAA1199 mRNA expression to the lower portion of normal epithelial crypts (A and B) and revealed that expression is markedly up-regulated in colorectal tumors (C). Asterisk, note the different levels of expression in tumor glands and normal crypts. D. KIAA1199 protein expression is also limited to the lower half of the normal colonic crypts, and a similar pattern is observed in the ileal mucosa (E), where the protein is expressed only in the crypts (not in the villi). In F and G, adenomatous crypts with low-grade dysplasia present increased expression of KIAA1199, particularly in the cytoplasm facing the crypt lumen, and in and around the mucin vacuoles of goblet cells (note the striking difference with goblet cells of normal crypts in both panels). The expression pattern changes dramatically during the transition from low-grade dysplasia with goblet cell differentiation (H) to high-grade dysplasia in which this differentiation is no longer apparent. J. In more advanced colon tumors, KIAA1199 overexpression is maintained. Note that, in I and J, the expression of KIAA1199 protein (like that of KIAA1199 mRNA; C) is highest in the luminal portion of the dysplastic glands (arrowheads, multilayer of unstained nuclei occupying more than the basal half of the dysplastic epithelium). K. Normal mucosa, with the corresponding tumor in the inset. Negative control: KIAA1199 antibody preabsorbed with the peptide used to immunize rabbits.

>4 higher or lower than those observed in the normal mucosa, but there were also several smaller changes in this category (Supplementary Table S6) that might also have dramatic effects on gene expression. Several other alterations reported in Table 2 are noteworthy in terms of their potential effect on cell proliferation, differentiation, apoptosis, and cell adhesion: (a) up-regulation of four members of the REG (regenerating) family of genes (21, 22), which would lead to increased tissue mitogen expression; (b) up-regulation of LCN2 (23) and down-regulation of ZFHX1B/SIP-1 (24) in the absence of significant changes in the expression of the epithelial cadherin CDH1 (E-cadherin), which would prevent or delay the epithelial-mesenchymal transition [changes were also noted in the expression of other cell adhesion genes of the cadherin and claudin families, including the striking overexpression of the placental cadherin gene CDH3, which is associated with early events in the transformation process (25, 26)]; (c) down-regulation of ZFHX1B/SIP-1 and Max dimerization protein 1 (MXD1/MAD1; decreased only 3.3-fold and therefore not listed in Table 2; refs. 27, 28) and overexpression of the RTEL1 helicase, which should facilitate telomere elongation (29); (d) alterations that would diminish apoptosis [e.g., overexpression of the decoy receptor for Fas ligand, TNFRSF6B, which is reportedly coregulated with RTEL1 on chromosome 20q13.3

Table 2. Genes Most Likely to be Involved in the Development and Evolution of Colorectal Adenomas (A Subset of Genes Listed in Supplementary Table S4) Subdivided by Gene Ontology Category

Gene symbol	Gene name	Fold differences*	
		E	I
Regulation of transcription			
NLF1	Nuclear localized factor 1	33.1	
FOXQ1	Forkhead box Q1	24.4	
MSX2	Msh homeobox homologue 2	22.2	
ASCL2	Achaete-scute complex-like 2	17.3	
MSX1	Msh homeobox homologue 1	8.5	
IRX3	Iroquois homeobox protein 3	8.4	
GRHL3	Grainyhead-like 3	7.9	
TRIM29	Tripartite motif-containing 29	7.4	
ETV4	Ets variant gene 4 (E1A enhancer binding protein, E1AF)	5.4	
ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like 2	5.3	
TEAD4	TEA domain family member 4	5.2	
SP5	Sp5 transcription factor	5.2	
HES6	Hairy and enhancer of split 6	4.6	
TBX3	T-box 3	4.6	
NFE2L3	Nuclear factor (erythroid-derived 2)-like 3	4.3	
GRHL1	Grainyhead-like 1	4.2	
FEV	FEV (ETS oncogene family)		15.1
SPIB	Spi-B transcription factor		13.2
NEUROD1	Neurogenic differentiation 1		10.6
MEIS1	Meis1, myeloid ecotropic viral integration site 1		7.1
NR3C1	Nuclear receptor subfamily 3, group C, member 1		5.9
NR5A2	Nuclear receptor subfamily 5, group A, member 2		5.6
THRB	Thyroid hormone receptor, b		5.2
ZNF483	Zinc finger protein 483		5.1
ZFHX1B	Zinc finger homeobox 1b (SIP-1)		4.8
MEOX2	Mesenchyme homeobox 2		4.7
HOXD10	Homeobox D10		4.6
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene		4.5
SOX10	SRY (sex determining region Y)-box 10		4.2
Cell proliferation/differentiation/apoptosis			
REG1B	Regenerating islet-derived 1b	75.8	
REG3A	Regenerating islet-derived 3a	29.5	
TACSTD2	Tumor-associated calcium signal transducer 2	21.4	
IL-8	Interleukin-8	14.7	
SERPINB5	Serpin peptidase inhibitor, clade B, member 5 (Maspin)	13.8	
REG1A	Regenerating islet-derived 1a	8.2	
FAIM2	Fas apoptotic inhibitory molecule 2	7.5	
DUSP4	Dual specificity phosphatase 4	7.4	
REG4	Regenerating islet-derived family, member 4	6.8	
PHLDA1	Pleckstrin homology-like domain, family A, member 1	6.0	
LCN2	Lipocalin 2 (oncogene 24p3)	5.7	
RTEL1	Regulator of telomere elongation helicase 1	5.6	
TGFBI	Transforming growth factor, b induced	5.2	
IGFBP2	Insulin-like growth factor binding protein 2	4.8	
TDGF1	Teratocarcinoma-derived growth factor 1	4.7	
TNFRSF6B	Tumor necrosis factor receptor superfamily, member 6b, decoy	4.5	
DMBT1	Deleted in malignant brain tumors 1	4.2	
TNFRSF10C	Tumor necrosis factor receptor superfamily, member 10c, decoy	4.1	
ANGPTL1	Angiopoietin-like 1 (Angioarrestin)		24.9
CDKN2B	Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)		14.9
GPM6B	Glycoprotein M6B		11.5
ANK2	Ankyrin 2		9.8
UNC5C	Unc-5 homologue C		7.4
HPGD	Hydroxyprostaglandin dehydrogenase 15-(NAD)		6.1
CPNE8	Copine VIII		5.5
FAIM3	Fas apoptotic inhibitory molecule 3		5.4
IL6R	Interleukin-6 receptor		4.8
TUSC3	Tumor suppressor candidate 3		4.7
DUSP1	Dual specificity phosphatase 1		4.7
RERG	RAS-like, estrogen-regulated, growth inhibitor		4.6
NDN	Necdin		4.5
IGF1	Insulin-like growth factor 1 (somatomedin C)		4.0
Cell adhesion			
CDH3	Cadherin 3, type 1, P-cadherin	81.7	
CLDN2	Claudin 2	16.1	
CLDN1	Claudin 1	9.0	
DSG3	Desmoglein 3	7.3	

(Continued on the following page)

Table 2. Genes Most Likely to be Involved in the Development and Evolution of Colorectal Adenomas (A Subset of Genes Listed in in Supplementary Table S4) Subdivided by Gene Ontology Category (Cont'd)

Gene symbol	Gene name	Fold differences*	
		E	I
DSG4	Desmoglein 4	5.9	
CLDN8	Claudin 8		25.8
CDH19	Cadherin 19, type 2		8.3
CEACAM7	Carcinoembryonic antigen-related cell adhesion molecule 7		8.3
CLDN23	Claudin 23		8.0
NRXN1	Neurexin 1		7.1
PCDH19	Protocadherin 19		6.8
NLGN4X	Neuroligin 4, X-linked		6.0
TNXB	Tenascin XB		5.6
MUCDHL	Mucin and cadherin-like		5.1
PCDH9	Protocadherin 9		4.9
L1CAM	L1 cell adhesion molecule		4.2

*Overexpressed (E) or underexpressed (I) in adenomas (versus normal mucosa samples).

(30-32); decreased expression of the netrin-1 receptor, UNC5C (33); and expression changes involving three Fas apoptosis inhibitory molecules (FAIM), including FAIM1, which was increased 2.3-fold and is thus not listed in Table 2]; and (e) marked down-regulation of several genes that would result in reduced tumor suppression activity [e.g., those encoding the antiangiogenic factor ANGPTL1 (34), the cyclin-dependent kinase inhibitor CDKN2B/p15, and the prostaglandin catabolism enzyme HPGD (35)].

It is also important to recall the size-related differences noted in the adenoma gene expression profiles (Fig. 2; Supplementary Table S3). When validated in a larger series of tumors, these differences should provide important clues to the molecular basis of the well-known link between the dimensions and malignant potential of colorectal adenomas (1).

Our study also furnishes a complete picture of expression changes involving gene components of the Wnt pathway across the transition from normal to adenomatous epithelium (Supplementary Table S2) as well as evidence for the existence of a novel Wnt target: KIAA1199. This gene, which encodes a protein of unknown function, was strikingly overexpressed in all the adenomas included in this study and in 25 adenocarcinomas of the colon described in a previous report (8). Even more intriguingly, its expression was significantly correlated with that of several genes that are well-established targets of Wnt signaling. Our hypothesis that KIAA1199 is up-regulated by the TCF(s)/h-catenin transcription complex was considerably strengthened by the marked decreases in KIAA1199 expression observed in cultured colorectal cancer cells when the Wnt pathway was inhibited by overexpression of dominant-negative TCF4 proteins or by h-catenin knockdown. It is not yet clear whether this is a direct effect, but this possibility is supported by the results of a recent genome-wide TCF4 ChIP-on-chip analysis, which indicates that the KIAA1199 locus is surrounded by four TCF4-bound regions.[10] These findings are consistent with the probable role of this gene as a direct target of TCF4/h-catenin signaling in the intestine and in colorectal tumors.

Other features of KIAA1199 expression are also compatible with its putative role as a Wnt target gene. KIAA1199 mRNA and protein are both confined to the proliferative compartment of normal intestinal crypts, where Wnt signaling is normally active, and they are highly overexpressed in colorectal adenomas and carcinomas, where this pathway is almost always aberrantly activated.

In normal and tumor tissues, KIAA1199 is expressed in the cytoplasm of epithelial cells. In glands with low-degree dysplasia, higher concentrations are observed in the mucin vacuoles of goblet cells, but cytoplasmic expression of the protein in tumor cells remains elevated even after goblet cell differentiation has been lost (Fig. 4). These features, together with the localization of KIAA1199 in the luminal portion of the cytoplasm, are suggestive of a secreted and/or membrane protein. This conclusion is consistent with our in silico analysis of KIAA1199 (see Supplementary Data and Supplementary Fig. S5), which strongly predicts the presence of a signal peptide at its NH_2-terminal end. In addition, the central region of KIAA1199 contains a TMEM2 homology domain, which is present in several eukaryotic proteins, including TMEM2, polyductin (PKHD1), and fibrocystin L (PKHD1L1; Fig. 5), all large receptor proteins characterized by an NH_2-terminal signal peptide or a single transmembrane helix and a short cytoplasmic tail (36).

A study based on yeast two-hybrid screens suggested that KIAA1199 may interact with plexin A2 (KIAA0463; ref. 37). The transmembrane plexins interact with transmembrane semaphorins on nearby cells, providing "stop" and "go" signals that are crucial for cell motility and invasive growth (38, 39). KIAA1199/plexin A2 interaction could thus play important roles in colorectal tumorigenesis not only in the invasive stages but also earlier during the formation of abnormal glands in benign adenomas.

A recent report linked high levels of KIAA1199 mRNA with cell mortality in human fibroblasts and in a renal cell carcinoma cell line (40). In that study, however, there was no significant increase in KIAA1199 expression during replicative aging of mortal cells, and this finding contrasts with the documented behavior of other genes involved in cell aging (41). Furthermore,

[10] Hatzis et al., unpublished data.

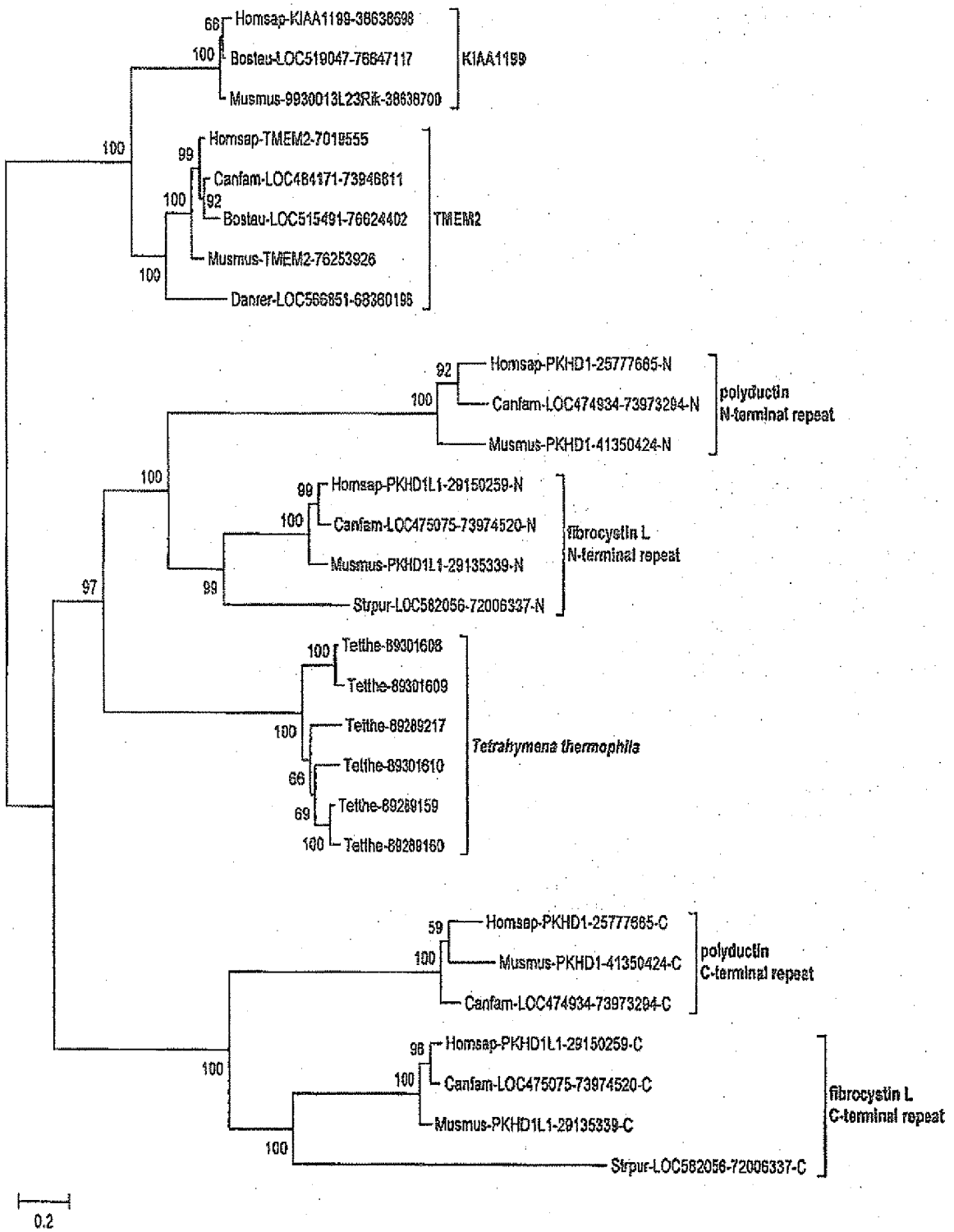


FIGURE 5. Phylogenetic tree of the proteins containing the TMEM2 homology domain found in the central region of KIAA1199. The tree was generated with MEGA3 (52) from the multiple sequence alignment shown in Supplementary Fig. S5. It was calculated with the minimum evolution algorithm and the JTT matrix. Positions with gaps were removed for calculation of pairwise distances. Node robustness was assessed using the bootstrap method with 100 resamplings. (Bootstrap values are shown at the nodes.) Two branches emerged, one comprising KIAA1199 and TMEM2 and the other with polyductin, fibrocystin L, and several other THD-containing proteins found in the ciliate Tetrahymena thermophila, which were apparently generated in a series of Tetrahymena-specific gene duplications. The NH_2-terminal repeats of polyductin and fibrocystin L clustered together, as did the COOH-terminal repeats, suggesting that the intragenic duplication of the TH domain in the ancestor of polyductin and fibrocystin L occurred before the divergence of chordates and echinoderms (more details in Supplementary Data).

the authors reported wide variation in KIAA1199 mRNA expression in breast cancer cell lines, and this finding raises the possibility that expression of this gene in vivo and in cell lines may differ.

We believe that our microarray data will serve as a springboard and reference point for other studies on the molecular basis of colorectal transformation along the adenoma-carcinoma pathway (and subsequently for the study of alternative pathways). Some of the transcriptional changes reported in this study might one day be used as molecular indices of the susceptibility of adenomas to malignant transformation, information that would be helpful in planning appropriate follow-up of the lesions. As for KIAA1199, its invariably high expression in the colorectal tumors we studied raises interesting possibilities for the development of a new molecular marker for the detection of these neoplasms. For example, because KIAA1199 expression in the normal mucosa is limited to cells in the lower portion of the crypts, which are not yet programmed to be shed into the intestinal lumen, the presence of KIAA1199 peptides in fecal water might prove to be a specific marker of adenomatous lesions. In addition, although due consideration must be given to its probable physiologic role(s) in intestinal crypts and possibly in several other human tissues (40, 42, 43), KIAA1199 may be a potential target of antibody-based therapies.

Materials and Methods

Tumor Samples

Pedunculated colorectal polyps and normal mucosa were obtained during colonoscopies carried out in the Gastroenterology Unit of the Belcolle City Hospital (Viterbo, Italy). The tissues were collected prospectively with informed patient consent and the approval of the local Human Research Ethics Committee. Patients with documented familial polyposis, with >15 adenomatous polyps (total: synchronous + previously excised; ref. 44), or currently treated with nonsteroidal anti-inflammatory drugs (including aspirin) were excluded from the study.

For each polyp, three biopsies of normal mucosa were collected from the same colon segment (≥2 cm from the site of the polyp). Immediately after removal, a small sample of epithelial tissue (5-15 mg) was cut from the tip of each polyp, leaving the underlying muscularis mucosae intact. We excluded polyps <1 cm to ensure that the sampling procedure would not interfere with the histologic diagnosis. All polyp samples were collected by a single operator (M.d.P.) using the same procedure to minimize artifacts due to sampling differences. The approach used allowed us to obtain specimens with a high percentage of epithelial cells without resorting to microdissection, which can diminish the quantity and quality of the extracted RNA.

The polyp sample and the three normal mucosal biopsies were immersed in RNAlater (Ambion) for subsequent microarray analysis, and the remainder of the polyp was submitted for pathologic analysis. The cut surface at the tip was labeled with India ink so that the sampled area could be easily identified during routine histologic examination. The tissue was then fixed in buffered formalin and embedded in paraffin. DNA extracted from sections of this specimen was also used to rule out microsatellite instability (reflecting defective DNA mismatch repair) at the BAT26 locus, as previously described (45). All of the polyps included in the study met the following criteria: type 0-Ip (6), maximum diameter of 1 to 4 cm, absence of surface ulceration, histologic diagnosis of adenoma, and absence of microsatellite instability at BAT26.

In some analyses, we also included transcriptomic data from a previously described set of 25 colon cancers (mismatch repair proficient and deficient; ref. 8), which we reanalyzed for this study with the same microarray used to characterize the adenomas and normal mucosa.

Microarray Analysis, Real-time Reverse Transcription-PCR, and Northern Blotting

Total RNA was extracted (RNeasy Mini kit, Qiagen) from homogenized tissue samples (5-15 mg), and its integrity was verified by capillary gel electrophoresis (Bio Analyzer, Agilent Technologies). Complementary RNA (15 Ag/sample), synthesized and labeled as previously described (8, 46), was hybridized with the Affymetrix U133 Plus 2.0 array, which contains in situ synthesized oligonucleotides representing the entire human genome (54,675 probes).

Raw gene expression data generated by GeneChip Operating Software (Affymetrix) were imported into the GeneSpring software program (Agilent Technologies) and normalized per chip (i.e., to the median of all values on a given array) and per gene (i.e., to the median expression level of the given gene across all samples). Analysis was done using the log expression values with GeneSpring's cross-gene error model turned on. Probes were excluded from analysis unless they were listed as "present or marginal calls" and/or had expression values ≥100 in ≥50% (≥16 of 32) of the samples in at least one of the tissue groups (adenomas and normal mucosa).

Expression data were subjected to four different unsupervised analyses: (a) hierarchical clustering using the Pearson correlation coefficient as a similarity measure and the average linkage algorithm for branch merging; (b) PCA, which reduces the dimensionality (number of variables) of a data set while retaining most of its variance (8); (c) correlation analysis, which involved computation of Pearson correlation coefficients for all possible sample pairs and visualization of correlation values as tile plots; and (d) CA, another dimension-reducing method (47), which was used to identify samples associated with particular gene expression levels. In typical CA, a matrix of n gene expression levels from p samples is treated as a two-way contingency table (genes by samples or vice versa) with n and p specifications for the "factors" gene and sample, respectively. Each intensity value thus reflects the abundance of a given transcript in a given sample. Like PCA, CA identifies independent "factorial components" that account for variance within a multidimensional gene data set, but in this case, the components are identified and ranked according to the correlation between gene and sample scores. A supervised or constrained extension of CA (9), CCA, was then used to identify possible correlations between gene expression patterns and clinical or pathologic variables. CA and CCA, as well as the corresponding plots, were computed using R software and the ade4 and made4 packages furnished by Bioconductor.[11]

[11] http://www.bioconductor.org

The Mann-Whitney test was used to select genes differentially expressed in normal mucosa and adenomas; Benjamini-Hochberg multiple testing correction was applied with a false discovery rate of 0.01. The genes in this set that were differentially expressed with fold differences of ≥2.0 were then analyzed with ErmineJ software (48) to identify any biological processes from the Gene Ontology database (49) that were overrepresented.

Pearson correlation was used to identify correlation between KIAA1199 expression and the expression of other genes in the entire set of tissue samples. Fisher's exact test was used to identify possible overrepresentation of known Wnt targets among genes whose expression was closely correlated with that of KIAA1199 (correlation values ≥0.8).

Reverse transcription-PCR and Northern blotting were done as previously described (46, 50) to verify the expression level of KIAA1199 in tissue samples and in LS174T colon cancer cells in which inducible inhibition of the Wnt pathway had been achieved with previously described methods (14-16).

In situ Hybridization

Digoxigenin-labeled KIAA1199 antisense riboprobes were synthesized from a PCR product amplified from human colon cDNA with KIAA1199-specific primers (sense: 5[-cacatcggg-gaggagataga-3[; antisense, containing a T7 RNA polymerase-binding site: 5[-taatacgactcactatagggttccagacttgaca-3[). This product was transcribed in vitro using the DIG RNA labeling kit and T7 RNA polymerase (Roche Diagnostics). In situ hybridizations were done on paraffin-embedded sections of human colon fixed with 4% buffered formalin as described elsewhere (51).

Immunohistochemistry

Our in silico analysis of KIAA1199 (see Supplementary Data) indicated that residues 202 to 217 (IHSDRFDTYRSKKESE) form a loop between a conserved h-strand and the following helix of the NH_2-terminal GG domain. This charged, surface-exposed peptide was used to raise a rabbit polyclonal antibody, which was purified by affinity chromatography on Thiopropyl Sepharose 6B (Amersham) derivatized with the antigenic peptide. A 1:1,000 dilution of this antibody was used, as previously described (45), to evaluate KIAA1199 expression in formalin-fixed, paraffin-embedded sections of adenoma and normal mucosal tissues.

Acknowledgments

We thank the patients who participated in this study; R. Haider, E. Pani, T. Lehmann, A. Patrignani, and U. Wagner for technical assistance; A. Mueller, E. Veljkovic, and F. Bennwart for critical reading of the manuscript; and M. Kent for editorial assistance.

References

1. Winawer SJ, Zauber AG, Fletcher RH, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. Gastroenterology 2006;130:1872–85.

2. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history of untreated colonic polyps. Gastroenterology 1987;93:1009–13.

3. Pinto D, Clevers H. Wnt, stem cells and cancer in the intestine. Biol Cell 2005; 97:185–98.

4. Korinek V, Barker N, Morin PJ, et al. Constitutive transcriptional activation by a h-catenin-Tcf complex in APC−/− colon carcinoma. Science 1997;275: 1784–7.

5. Morin PJ, Sparks AB, Korinek V, et al. Activation of h-catenin-Tcf signaling in colon cancer by mutations in h-catenin or APC. Science 1997;275:1787–90.

6. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy 2005;37:570–8.

7. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 2007;50:113–30.

8. di Pietro M, Sabates Bellver J, Menigatti M, et al. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. Gastroenterology 2005;129:1047–59.

9. Ter Braak C. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 1986;67:1167–79.

10. Miki R, Kadota K, Bono H, et al. Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. Proc Natl Acad Sci U S A 2001;98:2199–204.

11. Wu LF, Hughes TR, Davierwala AP, Robinson MD, Stoughton R, Altschuler SJ. Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. Nat Genet 2002;31:255–65.

12. Zhang W, Morris QD, Chang R, et al. The functional landscape of mouse gene expression. J Biol 2004;3:21.

13. Van der Flier LG, Sabates-Bellver J, Oving I, et al. The intestinal Wnt/TCF signature. Gastroenterology 2007;132:628–32.

14. van de Wetering M, Barker N, Harkes IC, et al. Mutant E-cadherin breast cancer cells do not display constitutive Wnt signaling. Cancer Res 2001;61: 278–84.

15. van de Wetering M, Sancho E, Verweij C, et al. The h-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002;111:241–50.

16. van de Wetering M, Oving I, Muncan V, et al. Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. EMBO Rep 2003;4:609–15.

17. Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 2001;61:3124–30.

18. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 2002;21: 4120–8.

19. Nosho K, Yamamoto H, Adachi Y, Endo T, Hinoda Y, Imai K. Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis. Br J Cancer 2005;92:1193–200.

20. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F. Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia. Gut 2003;52:1148–53.

21. Nata K, Liu Y, Xu L, et al. Molecular cloning, expression and chromosomal localization of a novel human REG family gene, REG III. Gene 2004;340:161–70.

22. Bishnupuri KS, Luo Q, Murmu N, Houchen CW, Anant S, Dieckgraefe BK. Reg IV activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas. Gastroenterology 2006;130:137–49.

23. Hanai J, Mammoto T, Seth P, et al. Lipocalin 2 diminishes invasiveness and metastasis of Ras-transformed cells. J Biol Chem 2005;280:13641–7.

24. Comijn J, Berx G, Vermassen P, et al. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol Cell 2001;7:1267–78.

25. Jenkowski JA, Bedford FK, Boulton RA, et al. Alterations in classical cadherins associated with progression in ulcerative and Crohn's colitis. Lab Invest 1998;78:1155–67.

26. Hardy RG, Tselepis C, Hoyland J, et al. Aberrant P-cadherin expression is an early event in hyperplastic and dysplastic transformation in the colon. Gut 2002; 50:513–9.

27. Lin SY, Elledge SJ. Multiple tumor suppressor pathways negatively regulate telomerase. Cell 2003;113:881–9.

28. Mariadason JM, Nicholas C, L'Italien KE, et al. Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. Gastroenterology 2005;128:1081–8.

29. Ding H, Schertzer M, Wu X, et al. Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. Cell 2004;117:873–86.

30. Pitti RM, Marsters SA, Lawrence DA, et al. Genomic amplification of a

decoy receptor for Fas ligand in lung and colon cancer. Nature 1998;396:699–703.

31. Bat C, Connolly B, Metzker ML, et al. Overexpression of M68/DcR3 in human gastrointestinal tract tumors independent of gene amplification and its location in a four-gene cluster. Proc Natl Acad Sci U S A 2000;97:1230–5.

32. Postma C, Hermsen MA, Coffa J, et al. Chromosomal instability in flat adenomas and carcinomas of the colon. J Pathol 2005;205:514–21.

33. Thiebault K, Mazelin L, Pays L, et al. The netrin-1 receptors UNC5H are putative tumor suppressors controlling cell death commitment. Proc Natl Acad Sci U S A 2003;100:4173–8.

34. Dhanabal M, Jeffers M, LaRochelle WJ, Lichenstein HS. Angioarrestin: a unique angiopoietin-related protein with anti-angiogenic properties. Biochem Biophys Res Commun 2005;333:308–15.

35. Backlund MG, Mann JR, Holla VR, et al. 15-Hydroxyprostaglandin dehydrogenase is down-regulated in colorectal cancer. J Biol Chem 2005;280:3217–23.

36. Hogan MC, Griffin MD, Rossetti S, Torres VE, Ward CJ, Harris PC. PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression. Hum Mol Genet 2003;12:685–98.

37. Nakayama M, Kikuno R, Ohara O. Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs. Genome Res 2002;12:1773–84.

38. Tamagnone L, Artigiani S, Chen H, et al. Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. Cell 1999;99:71–80.

39. Comoglio PM, Tamagnone L, Giordano S. Invasive growth: a two-way street for semaphorin signaling. Nat Cell Biol 2004;6:1155–7.

40. Michishita E, Garces G, Barrett JC, Horikawa I. Upregulation of the KIAA1199 gene is associated with cellular mortality. Cancer Lett 2006;239:71–7.

41. Horikawa I, Parker ES, Solomon GG, Barrett JC. Upregulation of the gene encoding a cytoplasmic dynein intermediate chain in senescent human cells. J Cell Biochem 2001;82:415–21.

42. Abe S, Katagiri T, Saito-Hisaminato A, et al. Identification of CRYM as a candidate responsible for nonsyndromic deafness, through cDNA microarray analysis of human cochlear and vestibular tissues. Am J Hum Genet 2003;72:73–82.

43. Abe S, Usami S, Nakamura Y. Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in Deiters' cells and the fibrocytes, as the cause of nonsyndromic hearing loss. J Hum Genet 2003;48:564–70.

44. Marra G, Jiricny J. Multiple colorectal adenomas—is their number up? N Engl J Med 2003;348:845–7.

45. Truninger K, Menigatti M, Luz J, et al. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. Gastroenterology 2005;128:1160–71.

46. di Pietro M, Marra G, Cejka P, et al. Mismatch repair-dependent transcriptome changes in human cells treated with the methylating agent N-methyl-N'-nitro-N-nitrosoguanidine. Cancer Res 2003;63:8158–66.

47. Benzécri J-P. Correspondence analysis handbook. New York: Marcel Dekker, Inc.; 1992.

48. Lee HK, Braynen W, Keshav K, Pavlidis P. ErmineJ: tool for functional analysis of gene expression data sets. BMC Bioinformatics 2005;6:269.

49. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.

50. Barker N, Hurlstone A, Musisi H, Miles A, Bienz M, Clevers H. The chromatin remodelling factor Brg-1 interacts with β-catenin to promote target gene activation. EMBO J 2001;20:4935–43.

51. Maake C, Auf der Maur F, Jovanovic K, Reinecke M, Hauri D, John H. Occurrence and localization of uroguanylin in the aging human prostate. Histochem Cell Biol 2003;119:69–76.

52. Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 2004;5:150–63.